INTRODUCTION TO INFRARED AND ELECTRO-OPTICAL SYSTEMS

SECOND EDITION

Ronald G. Driggers • Melvin H. Friedman • Jonathan M. Nichols

Introduction to Infrared and Electro-Optical Systems

Second Edition

For a listing of recent titles in the *Artech Optoelectronics and Applied Optics Series*, turn to the back of this book.

Introduction to Infrared and Electro-Optical Systems

Second Edition

Ronald G. Driggers
Melvin H. Friedman
Jonathan Nichols



ARTECH
HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalog record for this book is available from the British Library.

ISBN-13: 978-1-60807-100-5

Cover design by Vicki Kane

© 2012 Artech House

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

10 9 8 7 6 5 4 3 2 1

This book is dedicated to Bernard and Phillip Friedman who gave direction and purpose to my life when I was a young boy. You are loved and will never be forgotten.

Your loving brother,
Mel

It is also dedicated to the memory of Jill McGregor who picked me up when I was down and made me want to be a better person.

Your forever boyfriend,
Ron

Contents

Preface *xiii*

CHAPTER 1
Introduction 1

1.1 Introduction to Imaging 2
1.2 Infrared and EO Systems 3
1.3 Wavelength Dependencies 4
1.4 Typical EO Scenario 6
1.5 Typical Infrared Scenario 7
1.6 Analytical Parameters 8
1.7 Sensitivity and Resolution 9
1.8 Linear Systems Approach 10
1.9 Summary 12
1.10 Guide to the References 13
References 13

CHAPTER 2
Mathematics 15

2.1 Complex Functions 15
2.2 Common One-Dimensional Functions 17
2.3 Two-Dimensional Functions 20
2.4 Convolution and Correlation 22
2.5 The Fourier Transform 27
2.6 Properties of the Fourier Transform 29
2.7 Transform Pairs 30
2.8 Probability 33
2.9 Important Examples 37
2.10 Guide to the References 39
2.11 Exercises 39
References 42
Software 42

CHAPTER 3
Linear Shift-Invariant Systems 45

3.1 Linear Systems 47
3.2 Shift Invariance 48
3.3 Basics of LSI Systems 48
3.4 Impulse Response 50

3.5 Transfer Function 55
3.6 System PSF and MTF Versus Component PSF and MTF 58
3.7 Spatial Sampling 59
3.8 Spatial Sampling and Resolution 62
3.9 Sampled Imaging Systems 64
3.10 Guide to the References 69
3.11 Exercises 69
References 71

CHAPTER 4
Diffraction 73

4.1 Electromagnetic Waves 74
4.2 Coherence 77
4.3 Fresnel and Fraunhofer Diffraction from an Aperture 82
4.4 Fraunhofer Diffraction from a Thin Lens 87
4.5 Thin Lens Optical System Diffraction PSF 88
4.6 Thin Lens Diffraction MTF 91
4.7 Calculating Diffraction MTF with Pencil and Paper 96
4.8 Programs for Calculating Incoherent Diffraction MTF 97
4.9 Applications of Diffraction Theory 108
4.10 Exercises 113
References 116

CHAPTER 5
Sources of Radiation 117

5.1 Radiometry and Photometry 118
5.2 Infrared Targets and Backgrounds 125
5.3 Electro-Optical Targets and Backgrounds 135
5.4 Other Sensitivity Considerations 142
5.5 Target and Background Spatial Characteristics 143
5.6 Typical Midwave and Longwave Contrasts and Solar Effects 152
5.7 Exercises 159
References 161

CHAPTER 6
Atmospherics 163

6.1 Atmospheric Components and Structure 163
6.2 Atmospheric Transmission 166
6.3 Absorption 168
6.4 Scattering 170
6.5 Path Radiance 173
6.6 Turbulence 173
6.7 Atmospheric MTF 177
6.8 Models 179
6.9 Model Discussion 181

6.10 Some Practical Considerations 183
6.11 Exercises 186
References 186

CHAPTER 7
Optics 189

7.1 Light Representation and the Optical Path Length 189
7.2 Reflection and Snell's Law of Refraction 191
7.3 The Thin Lens Ray-Tracing Rules and Gauss's Equation 193
7.4 Spherical Mirrors 203
7.5 Modeling the Thick Lens 205
7.6 Vergence 208
7.7 Multiple-Lens Systems 213
7.8 Field of View 215
7.9 Resolution 218
7.10 Aperture Stop, Pupils, and Rays 221
7.11 The f-Number and Numerical Aperture 224
7.12 Telescopes and Angular Magnification 236
7.13 Modulation Transfer Function 245
7.14 Aberrations 250
7.15 Optical Materials 252
7.16 Cold Stop and Cold Shield 253
7.17 A Typical Optical System 253
7.18 Diffraction Blur 256
7.19 Guide To the References 258
7.20 Exercises 259
References 262

CHAPTER 8
Detectors 265

8.1 Types of Detectors 265
8.2 Photon Detectors 267
8.3 Thermal Detectors 272
8.4 Charge-Coupled Devices 275
8.5 Detector Responsivity 276
8.6 Detector Sensitivity 278
8.7 Detector Angular Subtense 285
8.8 Scanning Configurations and Implementations 287
8.9 Detector Transfer Functions 292
8.10 Infrared Detectors 296
8.11 Electro-Optical Systems 298
8.12 Noise 299
8.13 Basic Background-Limited Infrared Photodetection 300
8.14 New Infrared Detector Arrays 301
8.15 Exercises 305
References 306

CHAPTER 9
Electronics 309

9.1 Detector Circuits 309
9.2 Conversion of Spatial and Temporal Frequencies 313
9.3 Electronics Transfer Function 315
9.4 Noise 318
9.5 MTF Boost Filter 322
9.6 EO Mux MTF 322
9.7 Digital Filter MTF 324
9.8 CCDs 325
9.9 Uniformity Correction or "NUC" 326
9.10 Readout Integrated Circuits 328
9.11 Exercises 331
References 332

CHAPTER 10
Image Processing 333

10.1 Basics of Sampling Theory 333
10.2 Applications of Image Filtering 336
10.3 Super-Resolution Image Reconstruction 339
10.4 Image Fusion 346
10.5 Summary 350
References 351

CHAPTER 11
Displays, Human Perception, and Automatic Target Recognizers 355

11.1 Displays 355
11.2 Cathode-Ray Tubes 357
11.3 Light-Emitting Diodes 359
11.4 Liquid-Crystal Displays 361
11.5 Plasma Displays 363
11.6 Sampling and Display Processing 363
11.7 Human Perception and the Human Eye 365
11.8 Modulation Transfer Function of the Eye 367
11.9 Contrast Threshold Function of the Eye 369
11.10 Automatic Target Recognition 369
11.11 Exercises 371
References 372

CHAPTER 12
Historical Performance Models 373

12.1 Introduction 373
12.2 Johnson Model Fundamentals 374

12.3 The MRT Model 376
12.4 The First FLIRS and Models 378
12.5 Model Improvements for Resolution and Noise 381
12.6 Incorporating Eye Contrast Limitations 385
12.7 Model Improvement to Add Sampling 387
12.8 Other Improvements Prior to the Target Task Performance Metric 390
12.9 The TRM3 Model 391
12.10 Triangle Orientation Discrimination 392
12.11 Imager Modeling, Measurement, and Field Performance 393
12.12 Exercises 393
References 394

CHAPTER 13
Contrast Threshold and TTP Metric 397

13.1 Contrast Threshold Function of the Naked Eye 397
13.2 Contrast Threshold Function for the Eye–Display System 401
13.3 Validation of Eye–Display Contrast Threshold Model 410
13.4 Eye–Display Contrast Threshold Model 421
13.5 TTP Metric and Range Performance Modeling 426
13.6 Guide to the References 431
13.7 Exercises 431
Appendix 13A: Direct Calculations of $CTF_{eye-disp,h}$ 433
References 436

CHAPTER 14
Infrared and EO System Performance and Target Acquisition 439

14.1 Sensitivity and Resolution 442
14.2 Noise Equivalent Temperature Difference 443
14.3 EO Noise and Noise Equivalent Irradiance 449
14.4 Three-Dimensional Noise 455
14.5 Modulation Transfer Function 457
14.6 Minimum Resolvable Temperature Difference (Including 2D MRT) 459
14.7 Target Acquisition with Limiting Frequency (Johnson's N_{50}) 470
14.8 System Contrast Threshold Function 475
14.9 Target Acquisition with the Target Task Performance Metric (and Vollmerhausen's V_{50}) 480
14.10 Target Sets 482
14.11 Classic ISR, NIIRS, and General Image Quality 484
14.12 The Newest Military Imaging Mode: Persistent Surveillance 489
14.13 Exercises 491
References 491

CHAPTER 15

Search 493

15.1 Problem Definition 494
15.2 Introduction to Search Theory 495
15.3 Technique for Estimating Search Parameters and Their Uncertainties 499
15.4 Search Parameters and NV-IPM 503
15.5 Time-Limited Search 505
15.6 Field of Regard Search 510
15.7 Multiple Observers, Single Sensor, Unlimited Time, and Shared Knowledge 514
15.8 Independent Search with Two Sensors, Unlimited Time, and Shared Knowledge 517
15.9 Time-Dependent Search Parameters Search Model 518
15.10 Other Work 527
15.11 Guide to the References 530
15.12 Exercises 530
Appendix 15A: Time-Unlimited Field of Regard Search 531
Appendix 15B: Detection Times And Probabilities With Shared Information 534
Appendix 15C: Mathematica Search Code for TDSP Search Model 540
References 542

CHAPTER 16

Laboratory Measurements of Infrared Imaging System Performance 545

16.1 Sensitivity 545
16.2 Resolution 549
16.3 Human Performance: Minimum Resolvable Temperature Difference 551
16.4 Dynamic Minimum Resolvable Temperature Difference 553
References 555

List of Symbols 557
List of Acronyms 559
Table of Abbreviations and Concepts 561
Table of Operators and Mathematical Functions 563
Index 565

Preface

The second edition of this book is still intended to provide the reader with a complete introduction to infrared and electro-optical imaging systems. The first edition quickly became outdated and we are sure this will be the case with this edition as technology continues to progress at ever faster rates. The major changes from the first edition are that now systems can be characterized by a system contrast threshold function just like the human eye. Second, the target task performance metric can provide more accurate performance predictions than the previous limiting frequency methods. We added a section on target search and a section on laboratory testing of imagers. Finally, we included a summary of past performance modeling efforts that is useful to keep us ail grounded and remind us how system performance has evolved over the years.

This book includes a strong emphasis on the analysis and design of military imaging systems. Imaging systems can be analyzed with the same tools and techniques as electrical circuits and modern communications systems. The requisite courses for an undergraduate electrical engineering degree provide the majority of skills necessary to perform detailed system-level performance evaluations. Linear-shift-invariant (LSI) system principles and analytical function manipulations combined with calculus and domain transformations provide the mathematical foundations of imaging system analysis. Here, the system begins with the various sources of light, continues with its propagation through the atmosphere, the formation of an image, conversion to an electrical signal at the detectors, subsequent signal processing, and display to a human observer. This book presents a systems analysis approach, which allows an understanding of this complex process and results in quantitative characterizations of performance metrics such as modulation transfer functions, minimum resolvable temperature difference, and probability of object discrimination. We then use these metrics to analyze a number of imaging systems.

We begin with an introduction to two-dimensional functions and mathematics used to describe image transfer characteristics and imaging system component models. Next, LSI principles are presented to show image transfer characteristics for application to imaging systems. In this edition, we also describe work on how to address non-shift-invariant systems that are undersampled. The final background chapter is that of diffraction, showing the fundamental limits of imaging system performance. In this chapter, the diffraction concepts of coherent imaging systems are presented to show the development of the incoherent diffraction principles. This background information is included in Chapters 1 through 4.

Following the review of basic background material, the components of an imaging system are presented in Chapters 5 through 11. An imaging system "map" is used to chart the flow of information through the system. The map begins with the source of radiation and traverses through the atmosphere and into the imaging system. The path continues with the image forming optics and then onto the detector. The detector converts the light into an electrical signal that is amplified and processed by system

electronics. This electrical signal continues to a display or, in some cases, an automated or signal processor. A display converts the electrical signal into an optical signal for human visual consumption. The human eye is typically included as part of system performance, so it, too, is considered a component in the imaging system. Each of the components in the map is addressed in a chapter. These chapters include component descriptions, component analysis techniques, and LSI modeling descriptions.

The emphasis then turns to infrared and electro-optical system performance descriptions, design rules, and examples and systems issues. Chapters 12 through 16 provide infrared and electro-optical system performance through a historical performance model perspective, the contrast threshold of the eye, imaging system performance as calculated today, target search and detection, and laboratory testing of imagers. These are the chapters where the components are brought together to provide an overall system description. The sources of radiation and the atmosphere are typically not included in imaging system performance estimates intended for generic application to a variety of target, background, and atmospheric conditions. The analysis process then continues here as system-scenario performance, which includes the source, atmosphere, and imaging system performance characteristics to give discrimination performance of an imaging system within a given scenario. These chapters provide a number of analysis-and-design exercises that solidify the LSI concepts.

There are two target audiences for this book. The first is the senior or first-year graduate student with an interest in electronic imaging systems. This book is intended to provide these students with the necessary tools to enter an engineering position in the field of infrared or electro-optical systems. It provides them with the nomenclature, component descriptions, and performance parameters to understand sensor-related issues. The second audience is the practicing engineer who requires a reference of sensor and basic scenario performance calculations. Numerous analyses and designs are given throughout the text.

The authors would like to thank a number of people who made this book possible. First, we would like to thank everyone who helped with the first book: Keith Krapels, Rich Vollmerhausen, Nancy Davis, Larry Andrews, Pierre Schonbaum, Kenneth Barnard, Michael Currin, Roy Williams, Carl Halford, Stacey Taylor, Raymond Deep, Glenn Vinson, John Schroeder, Norm Kopeika, Barry Johnson, John Gunderson, Glenn Boreman, Charles Bray, Rich Vollmerhausen, Gary O'Brien, Mike Lloyd, Shapour Ahmadi, Bill Blecha, John Leachtenaur, Lew Boyleston, Gerald Hoist, and Sarah Edwards. I would also like to provide a special thanks to the first edition co-authors Paul Cox and Tim Edwards as around half of the information in this book remains the same, I would like to thank all those who helped make a significant difference in this second edition including Jon Fanning, Jim Waterman, Mel Kruer, Mark Greiner, Paul Norton, Steve Moyer, and Todd DuBosq. Special thanks goes to Brad Preece and Joe Reynolds for their significant help and their successful attempt to keep us straight in an ever increasing complexity of systems and to Rich Vollmerhausen who changed the way systems are evaluated and designed (it was a joy to work with him during this change). Finally, I want to thank Keith Krapels for starting this book in a pub when I was weak and temporarily forgot how much work books are and especially Mel Friedman and Jon Nichols, my co-authors, who started and finished strong while keeping my spirits up!

Ron Driggers

CHAPTER 1

Introduction

In the past decade since the first edition of this book was written, significant progress has been made on how infrared and electro-optical systems are analyzed and designed. One new performance metric is the target task performance (TTP) metric; not to be confused with the target transfer probability function (TTPF). The TTP metric replaces the previous Johnson criteria approach and provides good analysis and design results for all imagers, including staring thermal imagers. It improves end-to-end system performance modeling by more accurately reflecting human perception.

Two large changes occurred around the turn of the century: (1) Sensors went from producing limited dynamic range analog outputs to digital 14- to 16-bit outputs. This change was critical for taking advantage of the increased sensitivity of staring focal plane arrays over the previous generation of scanned arrays. (2) This increased output also enabled the use of real-time image processing and conditioning. Both of these advances resulted in more than doubling the range performance of imaging infrared sensors.

In the past decade, image performance has been positively impacted by advances in signal processing. Examples of such signal processing include localized contrast enhancement, super-resolution, electronic zoom, and image fusion. These enhancements are important for gaining higher performance in an imaging system. In the case of image fusion, it is unclear how to model the change in range performance. We provide brief descriptions of the signal processing enhancements just mentioned, but leave it to other sources to provide detailed descriptions of how advances in signal processing are approached in terms of modeling range performance.

Significant advances have also been made in detector technology. It is now becoming common to have midwave and long-wave infrared capabilities on the same focal plane. This combined capability on the same focal plane is called a *dual-band infrared system* and is considered by the U.S. Army to be the "third generation" of thermal imagers. The combined bands give an overall higher performance because the higher performing band can be used in any given situation (e.g., cold weather, smoke, wet targets, long-range identification). Dual-band technology and other advances in detector technology are covered in this book.

It is an explosive time in infrared and electro-optical system technology development. New imaging system capabilities and methods are evolving at an amazing rate. Areas of recent imaging technology emphasis are laser range-gated imaging and short-wave infrared systems. New detector technologies include type II superlattice and large-format persistent surveillance focal planes. These are only a few examples and the new technology keeps coming. We expect this book to be of limited use in another ten years due to this rate of technology development. However, we attempt to cover the basics and provide the reader with a timely learning resource and reference.

1.1 Introduction to Imaging

Humans have five senses: sight, sound, touch, taste, and smell. The data rates associated with each of these senses vary considerably. Most scientists agree that touch, taste, and smell have extremely low information rates compared with those of sight and sound. Using generous estimates of a 12-bit sound amplitude resolution and a hearing bandwidth of 20 kHz, the human sound data rate is in the ballpark of 240 Kbps. A conservative estimate of the sight bandwidth with 8 bits of amplitude resolution, 125 million picture elements (or retinal detectors [1]), and a 10-Hz imaging rate results in 10 Gbps. These data rates confirm our commonsense appreciation of vision and verify that the visual data volume to our brain is orders of magnitude larger than all the other senses combined.

In likeness of the eye, humans have created a large number of vision, or imaging, devices with high-data-rate information collection capabilities. Imaging devices range from endoscopes used in surgery to night vision devices used on the battlefield. The analysis and design techniques presented in this book are applicable to these types of imagers.

The first major category of examples and scenarios used in this text concentrates on long-range (i.e., large object distance) imagers. A major design and analysis issue with these systems is whether a particular object can be discriminated at a given distance under defined conditions. The second major category of applications is best described as situational awareness sensors. These sensors are for driving, flying, and personal mobility. In this application, the task is not to discriminate a target from the background, but to discern details of the scene or background. The last major category is surveillance and threat warning. These applications tend to be wide field of view or regard and utilize fixed focal length sensors. They combine some of the characteristics of targeting sensor performance and some of the characteristics of situational awareness but with an increased emphasis on automated processing instead of human observers. The subject matter in this book provides both the designer and the analyst with the necessary tools to address these issues.

The title of this book, *Introduction to Infrared and Electro-Optical Systems*, may seem redundant because *electro-optics* has been previously defined as "the field of systems that convert photons to electrons," [2] regardless of wavelength. Systems that include the spectral region of 0.4- to 0.7-μm wavelengths, the visible light band, are described as *visible* sensors. Those responding to the spectral region (0.7- to 14-μm wavelength light) are referred to as *infrared* sensors. The infrared spectrum is further divided into four subregions. The near-infrared (NIR) region is from 0.7 to 1.1 μm, the short-wave infrared (SWIR) region is from 1.1 to 3 μm, the midwave infrared (MWIR) region is from 3 to 5 μm, and the long-wave infrared (LWIR) region is from 8 to 14 μm.

In this book, electro-optical (EO) systems are those that respond to wavelengths within the 0.4- to 3-μm region (note that this includes the visible, NIR, and SWIR bands). Typically, most of the light collected by an EO sensor corresponds to light that was reflected by objects in the scene. The 3- to 5-μm (MWIR) and the 8- to 14-μm (LWIR) band imagers are collectively infrared sensors and are called *forward-looking infrared (FLIR)* systems or *imaging infrared (I^2R) sensors*. Typically, these sensors respond primarily to light that is emitted by objects in the scene. Although the analytical

techniques for the FLIR and EO sensors are similar, the performance descriptors are different. For example, many of the infrared imager performance parameters are described with the concept of differential temperature quantities, where EO systems are described in terms of radiant contrast.

1.2 Infrared and EO Systems

Figure 1.1 is presented throughout this book as an example of a standard model (and road map) in the study of infrared imager and EO systems. The diagram depicts a cascaded system where the input signal is the optical flux associated with the target, and the output is an image presented by a display for human consumption. The object flux is either reflected external radiation or self-emitted radiation. This flux then passes through the atmosphere, which degrades the amplitude and changes the phase of the radiation. These effects manifest themselves as a reduction in information content of the signal. Furthermore, particles in the atmosphere reflect unwanted stray light into the sensor. The EO or infrared *sensor's* first component is the optics that collects the photons from the target, background, and stray atmospheric flux. Then, the optics image the scene onto a detector plane. The detector transforms the light into electrical signals that represent the spatial distribution of the flux amplitude leaving the scene. The detector system may be a *staring array,* a *linear array,* or a *single detector*. A staring array is a two-dimensional (2D) arrangement of detectors and a linear array is a column (or a few columns) of detectors. In a linear array system, a scanner moves the image across the detectors. For a single detector system, a 2D scanner moves the image across the detectors in the horizontal and vertical dimensions. Under the detector array is the readout integrated circuit or ROIC.

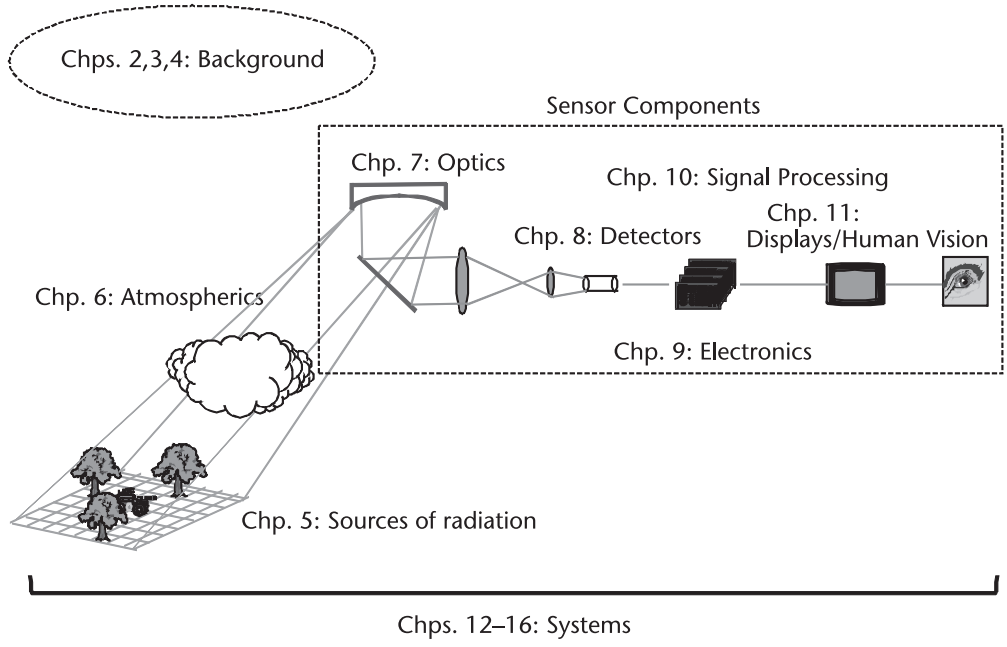


Figure 1.1 I^2R and EO systems.

Current ROIC technology is principally analog. An example is a charge coupled device with electron wells to store charge and a clock to read the wells out progressively by rows or columns. Research into first-generation digital ROICs is aimed at reducing the power and complexity of the ROICs. Electronics consist of preamplifiers, postamplifiers, filters, processors, digitizers, and digital signal processors. These components can be considered signal conditioners because the primary purpose of the electronics is to prepare the detector signal for either display or automated image processing. The electronic output signal can then be converted to an optical signal that can be displayed for human vision. In some systems, the signal undergoes additional analysis in a target tracker, automatic target recognizer (ATR), or an aided target recognizer (AiTR), whose output is used for operator cueing.

The goal of this book is to provide the reader with the analytical tools required to relate input to output by stepping through each of the above-described steps. In the process the reader will be made aware of the key system parameters and the tradeoffs involved in system design (e.g., sensitivity vs. resolution). We will also expose the reader to the latest technologies available for modern-day infrared and EO systems.

1.3 Wavelength Dependencies

The design or analysis of an imaging system for a given purpose yields performance indicators that are strong functions of wavelength. This spectral dependence results from many factors, including the scene's characteristics, atmospheric degradation, and the individual response of the sensor's components.

First, the object of interest, or target, has two surface characteristics defined as *reflectivity* and *emissivity*. Typically with EO sensors, the important target parameter is reflectivity, whereas emissivity is typically dominant in infrared sensor analysis. That is, targets with reflectivities near that of their backgrounds are difficult for EO and NIR sensors to exploit. Targets with emissivity (and temperature) characteristics near that of their backgrounds are difficult for infrared sensors to exploit. Both reflectivity and emissivity are functions of wavelength. A detailed discussion of these characteristics is presented in Chapter 5.

Once a photon leaves the target in the direction of the sensor, the atmosphere changes the photon path by refraction or scattering. The process whereby the atmosphere transmits or absorbs photons is highly dependent on the characteristic frequency or wavelength of the radiation. Figure 1.2 shows the atmospheric transmission (MODTRAN) of a 1-km horizontal path length at sea level under tropical and subarctic winter conditions. Tropical and subarctic winter conditions were chosen since they are expected to bracket transmission under normal conditions, for example, no smoke. Note that transmission is high for the visible, NIR, SWIR, MWIR, and LWIR bands. Sensors are not effective in the 5- to 8-μm band because of poor atmospheric transmission at these wavelengths.

The transmission of light through an imaging sensor can also be characterized as a function of wavelength. The optical sensor components, comprising lenses, filters, and mirrors, have reflectivity and transmission characteristics that are highly wavelength dependent. Finally, detectors convert photons to electrons with an efficiency that is a strong function of wavelength.

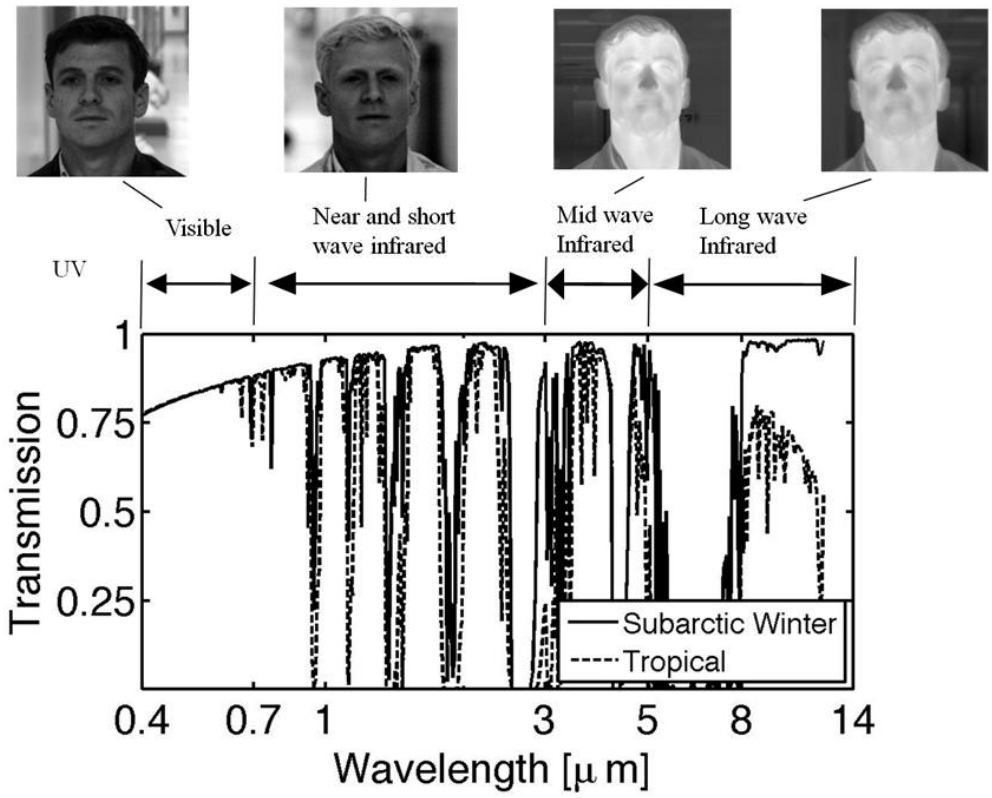


Figure 1.2 Atmospheric transmission for a 1-km horizontal path.

The diffractive nature of radiation depends on wavelength. Diffraction can be thought of as the curving and spreading of electromagnetic waves that propagate through space. This phenomenon is observed in the atmosphere and in vacuum and is a consequence of the wave nature of electromagnetic radiation. Short wavelengths tend to propagate in straighter lines; long wavelengths tend to curve more. The pertinent factor is the ratio of the wavelength compared to the characteristic size of the obstruction.

Diffraction limits the quality of an imaging sensor. Ideally a point source is imaged as a point. Because of diffraction a point source is typically imaged as a blur called the *point spread function*. The blur or point spread function typically has an approximately circular shape whose diameter is proportional to the wavelength and inversely proportional to the size of the obstruction. This implies that diffraction effects are most important for small apertures and long wavelengths.

Each of the characteristics introduced above is discussed in more detail in later chapters. All of these characteristics must be considered to effectively design or analyze sensors. For example, averaging transmission or some other wavelength-dependent characteristic such as detector spectral response over a spectral band and using these averages for each system component may cause significant error in performance estimates. While the error may be acceptable in some limited applications, a more accurate solution requires the wavelength response of each component to be treated separately. The collective system response is then determined by summing the

individual wavelength responses of the entire system (object, atmosphere, sensor, display, and observer).

1.4 Typical EO Scenario

We define EO systems as those that utilize the 0.4- to 3.0-μm waveband. The human eye is responsive to part of this waveband (i.e., 0.4 to 0.7 μm); therefore, EO systems provide images that are similar to normal human vision. Some EO systems extend into the near- or short-infrared spectral region, but can include the human visual band. The signal from targets, backgrounds, and clutter (objects that could be easily mistaken for targets) typically has a large reflective component in the EO wavelength band. Illumination, sometimes called *external radiation*, is provided by sunlight, moonlight, starlight, or artificial means. Targets and backgrounds reflect the external radiation as shown in Figure 1.3. Reflectivity differences in the 0.4- to 0.7-μm band are perceived as color by the human visual system. In the EO wavelength band, surface reflectivity is a strong function of wavelength. The radiation is not only reflected by the targets and background, it is also reflected, or scattered, by atmospheric, aerosol, smog, or smoke particles. Radiation reflected by targets and backgrounds must also be transmitted through the atmosphere before being collected by the EO sensor. Transmission lessons from the atmosphere have prompted the development of human-made obscurants. Smoke is a widely used battlefield cloaking obscurant intentionally introduced to inhibit radiation transmission. Scattering by smoke particles between the target and sensor is intended to make target-background contrast negligible.

The signal that enters the sensor is a mixture of target and background reflections that have been transmitted through the atmosphere along with atmospheric scattered radiation. This radiation enters the sensor aperture and some percentage (again, as a function of wavelength) is transmitted to the sensing elements. The sensing elements can be detectors, tubes, or image intensifiers. Image intensifiers or

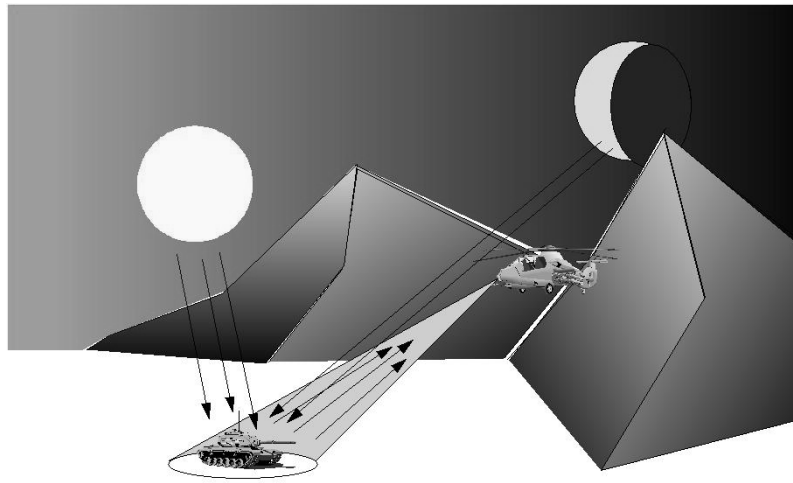


Figure 1.3 Typical EO scenario.

low-light detectors are used when moonlight, starlight, or a small active illuminator provides the source. The output of the sensing elements is amplified, processed by electronics, and routed to some monitor or display device. Humans then view the display device for information-gathering purposes.

A human uses the information shown on the display to detect, recognize, or identify targets. The detection task is the process of determining if a potential target is present. The recognition task is the gross classification of targets into major groups such as tanks, trucks, ships, cars, and the like. The identification task is a higher level of discrimination within a major group (e.g., discriminating a T-62 tank from a T-72 tank). In the past decade, a large number of discrimination tasks have been investigated so that metrics could be provided for these tasks. Force protection and antiterrorism activities have driven the development of these target metrics.

The output of the sensor electronics is not always displayed and viewed by a human for interpretation. In some systems, the output of the EO system is input to an ATR or AiTR. In these cases, a signal or image processor performs part or all of the detection, recognition, or identification tasks. In the AiTR case, the human bears the burden of the final classification level (i.e., an AiTR processor may detect objects, whereas the human performs the recognition or identification).

1.5 Typical Infrared Scenario

The typical infrared sensor provides a visual representation of an object at night and is especially useful when there is little visible light. The infrared sensor is used for battlefield night vision, surveillance of unlit areas, and fire detection within smoke-filled spaces onboard ships. There are two primary infrared transmission windows through the atmosphere: MWIR (3- to 5-μm band) and LWIR (8- to 14-μm band) and this convention is used by most engineers. In the remainder of the infrared band, the atmospheric absorption is too high for long-distance imaging (see Figure 1.2). MWIR and LWIR bands may mean slightly different spectral regions: sometimes the LWIR band is defined as the region between 8 and 12 μm.

A typical scenario for an infrared sensor is shown in Figure 1.4. Although solar reflections can influence the perceived image dramatically (especially in the MWIR), the usual mode of infrared operation is at night. All bodies above the temperature of absolute zero emit electromagnetic radiation according to Planck's law. A hotter object corresponds to higher levels of electromagnetic radiation. As the temperature of the object gets hotter, the peak wavelength moves to shorter wavelengths so that at very hot temperatures the radiation is perceived by the eye as light. The emissive surface characteristics of the hot object determine the spectral emission weighting of the radiation. The emitted radiation propagates through the atmosphere, where a percentage reaches the entrance aperture of the sensor. The amplitude and phase of the radiation entering the sensor aperture are altered by the transmission properties of the lenses and filters and by the reflectivity of the mirrors. When the radiation illuminates the detectors, a portion of the photons are converted into electronic signals. Like EO systems, these signals are presented to observers through displays or are input to computer target-recognition equipment.

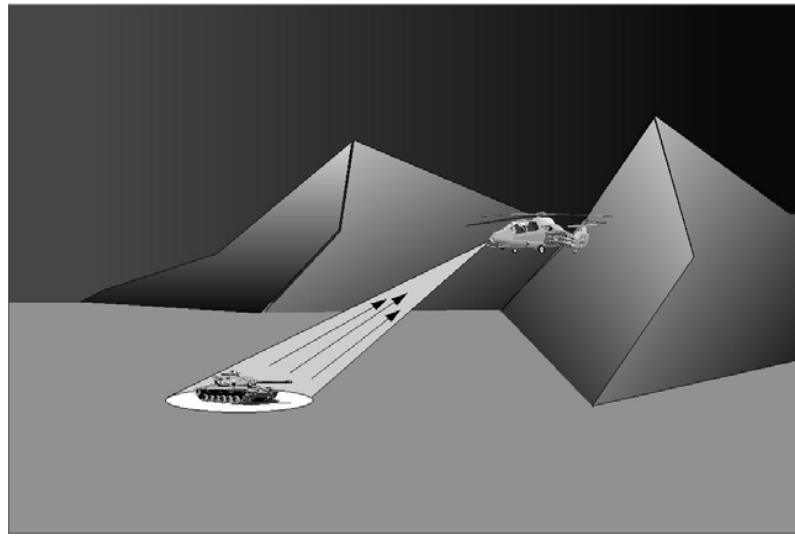


Figure 1.4 Typical infrared scenario.

Some engineers describe sensors that operate in the 3- to 5-μm or 8- to 12-μm wavebands as imaging infrared (I^2R) sensors but most describe these sensors as forward-looking infrared (FLIR) sensors. This historical term is also used to categorize the progression of infrared imaging through the designation of first-, second-, and third-generation FLIRs. The differences in the generations of FLIR technology are discussed in Chapter 12-16.

1.6 Analytical Parameters

Modeling of a typical infrared and EO scenario requires a large number of imaging system, target, and atmospheric parameters [3]. They can be grouped into illumination, target, atmospheric, optics, detectors, electronics, display or processing, and human visual parameters. Some of these parameters are listed in Table 1.1. Sensor analysis begins with a mathematical description of system components showing all independent variables. To make progress, approximations or simplifying assumptions are usually made. Many of the errors in sensor design are committed at this early stage. There is no substitute for a clear understanding of the required function of a system, characterization of the sensor components, and characterization of the target, background, and sources as well as atmospheric parameters.

Table 1.1 is not an exhaustive list of parameters. Many of them are functions of both wavelength and space. Some of the parameters change with time and some are relatively constant. Other important system requirements or attributes include mechanical specifications (jitter and drift), size, weight, power, cost, reliability, and maintainability.

One begins to see the complexity of infrared and EO sensor design. There are easily over 100 parameters to address before construction of a infrared or EO sensor can begin. Depending on the application, certain parameters may be more or less important. Without knowledge of the system's purpose or an understanding

Table 1.1 Analytical Parameters

Illumination	*Sensor detector*
Spectral irradiance	Responsivity
	Detectivity
Target	Noise characteristics
Reflectivity	Detector angular subtense
Emissivity	
Size and spatial characteristics	**Sensor electronics**
Temperature	Temporal characteristics
	Digital or analog
Atmosphere	filter characteristics
Weather conditions	
Obscurants	**Display**
Transmission	Resolution
Optical transfer function	Brightness and contrast
Scattering	
	Human psychophysics
Sensor optics	Temporal response
Lens transmissions	Image transfer function
Mirror reflectivity	Brightness dependence
Filter transmission	
Aberrations	**ATR or AiTR response**
Diffraction	Discrimination probability
Aperture size and shape	False-alarm rate

of the approximations and assumptions inherent in the model that describes the performance of the system, it is possible to inadvertently neglect one or more parameters and produce a system that fails to meet specifications.

1.7 Sensitivity and Resolution

The performance of infrared and EO sensors can be discussed in terms of two general parameters: *sensitivity* and *spatial resolution*. Sensitivity involves signal and noise levels including signals from the target, signals from the background, and noise signals that can make it difficult to see either the target or the background. Noise has a large variety of sources originating from both the sensor and the environment. Sensitivity is especially important when viewing objects with small image signals that can be drowned out by large noise sources. Frequently the image signal is the difference between the target and background signal. Having an adequate image signal is a major consideration when evaluating sensor designs. Radiometry is a discipline that describes the amount of radiation that leaves the object and eventually illuminates the sensor detector. Radiometry, optical design, detector performance characteristics, and detector sample times are of primary importance in sensor sensitivity analysis.

If a sensor has sufficient sensitivity, intelligent decisions on object detection, recognition, or identification can only take place with adequate resolution. Resolution is influenced by every stage of the process: the atmosphere, sensor optics, sensor detector, sensor electronics, and the display/observer. Important considerations are atmospheric turbulence, the diffraction or aberration limits of the optical system,

and the detector size and shape with respect to the optical focal length of the sensor. Electronic response, display resolution, and many other parameters also contribute to the resolution of the overall system.

In some cases, resolution and sensitivity are conflicting parameters. For example, for a well-designed system with a fixed optical diameter, an increase in the system focal length will typically increase resolution and decrease sensitivity and field of view. Sensitivity and resolution issues usually provide the most important sensor trade studies. Requirements analyses address the specification of a sensor that has sufficient sensitivity and resolution for a given application.

A comparison of EO and infrared sensors shows that sensitivity is more important for infrared systems. Detectors that respond to the visible and near-visible wavelengths typically provide better performance than those of infrared detectors. The shorter wavelength photons of the visible light provide a higher energy in the electron conversion process. Also, the development of visible detectors has been the subject of much research for a longer period of time. Therefore, the higher sensitivity of EO detectors and smaller size gives EO sensors better characteristics than infrared sensors. In general, EO systems have adequate sensitivity even with high resolution. Typical EO systems are resolution limited. Infrared systems are typically more sensitivity limited than EO sensors, so more attention must be paid to the sensitivity parameters in infrared system design. This is especially true for scanned infrared systems. Staring array infrared systems have much more signal to noise due to the amount of detector time on a target. Many infrared systems are background noise limited where other-than-target sources drown out the target signal (recall that all objects are emitting).

1.8 Linear Systems Approach

The linear systems approach to sensor analysis is convenient in determining the response of an imaging system. Imaging systems can usually be considered linear shift-invariant (LSI) systems, where the LSI system is defined as having the properties

$$L(af(x-c)+bg(x-d)) = aL(f(x-c)) + bL(g(x-d)) \tag{1.1}$$

where a and b are multiplicative constants and c and d are shifting constants. Here $f(x)$ and $g(x)$ are functions of the independent variable x, and L denotes the LSI operator.

Equation (1.1) states that a LSI system is one in which the response to a sum of different inputs can be obtained by considering the response to each input separately. The expression also states that shifting the independent variable does not change the result. The response of an imaging system can be considered the summation of a set of point-source responses. For example, the image of a point-source object provided by an EO or infrared system is desired to be a point-like image. Two point-source objects would produce two point-like images. The superposition aspect implies that the sum of two separately imaged point-source objects would produce the same two point-like images. Also, when an object is shifted, the image is shifted, but unchanged in appearance. Finally, a brighter point-source object would produce a brighter image.

Imaging systems can be treated as LSI systems in most cases, where any target or background source can be represented as an infinite number of weighted point radiation sources. The collective response of the sensor to these point sources yields an accurate representation of the sensor image. Again, there are two very important aspects of LSI sensor analysis: sensitivity and resolution.

Sensitivity is typically addressed through radiometric analysis. Superposition holds here as a characteristic of flux, or radiometric power (for noncoherent optical systems). Resolution is addressed in a manner very similar to that used in complex circuit and signal analyses. An impulse response is determined for the system and applied to the input signal by way of 2D convolution to produce an output image. This approach can also be performed in the spatial frequency domain. The imaging system transfer function is multiplied by the input object spectrum to produce the output image spectrum.

Each system contributor or component has its own impulse response and transfer function including the atmosphere, optics, detector, electronics, mechanical aspects, display, and human vision. The system impulse response can be determined by the convolution of all component impulse responses, and the system transfer function can be determined by the multiplication of all component transfer functions (just like in circuit analysis). Chapter 3 provides a detailed description of this approach.

While the LSI approach gives good results, not all components provide LSI responses. For example, the impulse response of the human eye is a nonlinear function of display brightness. These nonlinearities are treated by fixing parameters within a range such that the behavior of the component can be assumed linear.

A LSI sensor analysis can result in a characterization using several parameters. Sensitivity parameters are described by system intensity transfer functions (SITFs) and noise equivalent temperature differences (NETDs) or noise equivalent irradiance (NEI). Resolution performance can be given by an impulse response or a modulation transfer function (MTF) for the entire system. Sensitivity and resolution parameters can be combined to give an overall performance indication of minimum resolvable temperature differences (MRTD or just MRT) for infrared systems or minimum resolvable contrast (MRC) for EO systems. More recently, the sensitivity and resolution rollup is quantified by the system contrast threshold function (CTF). MRT, MRC, and CTF curves are provided as a function of spatial frequency (i.e., sensitivity plotted as a function of resolution).

The MRT, MRC, and CTF are important parameters because sensitivity and resolution are not independent. These functional curves give a sensor analyst a quantitative performance measure of a sensor in terms of both sensitivity and resolution.

A sensor user may want to know how far away a particular target can be detected, recognized, or identified. Because sensor performance curves are known, target size and contrast are needed along with viewing path atmospheric characteristics. A discrimination criterion is used to provide the minimum number of cycles across a target to detect, recognize, or identify the target. The spatial frequency corresponding to one of these tasks coupled to the atmospheric transmission and the target-to-background contrast can be used to determine the probability of detection, recognition, or identification of the target as a function of range (distance from the sensor to the target).

1.9 Summary

The material presented in this book is provided in three sections as shown in Figure 1.5. Part 1 provides a background review of the mathematics, LSI systems, and diffraction principles necessary to perform sensor analysis. Part 2 describes LSI system component modeling associated with the targets and backgrounds, atmospheric path, each of the sensor components, and the human visual response. Emphasis here is on introductory modeling of these components. Finally, Part 3 combines the LSI system methods and the component models to provide overall system-level performance estimates for infrared and EO sensors. Infrared systems are presented first because many of the principles that apply to EO systems were developed for infrared systems. Design issues are discussed where they affect sensor component performance. Intelligence, surveillance, and reconnaissance (ISR) systems are also covered along with persistent ISR systems (P-ISR). Examples of real systems are

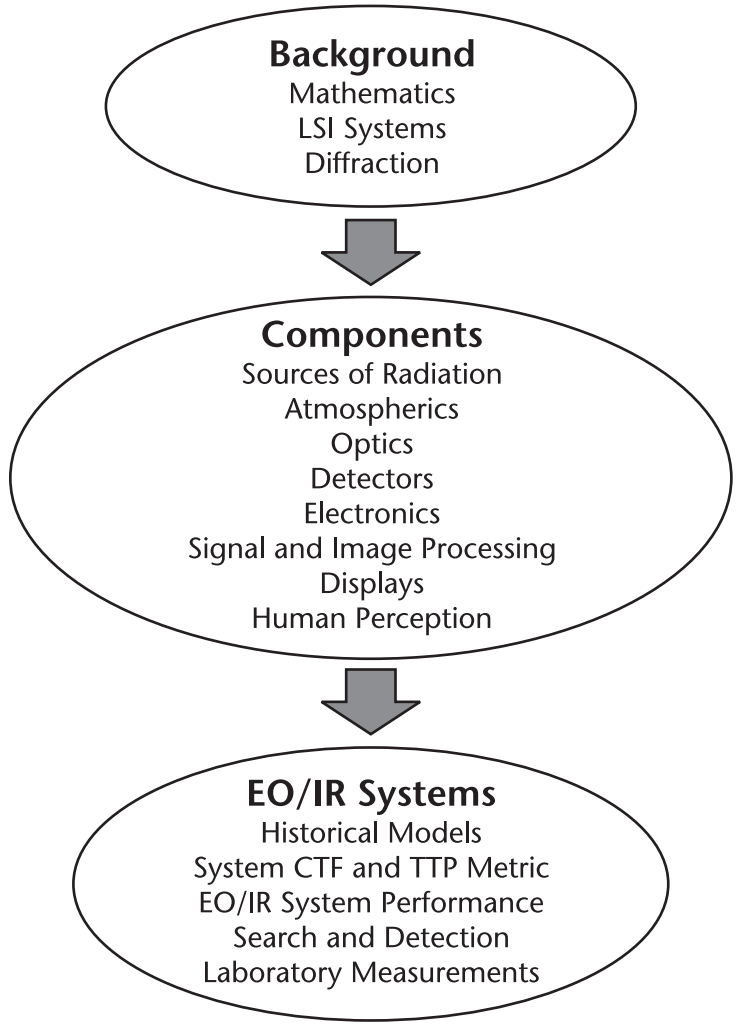


Figure 1.5 Organization of this book.

used throughout the text to illustrate analysis and design principles. Finally, we close out the book with some descriptions of new systems that are currently being fielded.

1.10 Guide to the References

No effort was made to give a comprehensive set of references. Instead we confined ourselves to those books we felt would be most useful to the reader.

References with content similar to that presented in this book include Daniels (2010) [4], Holst (2008) [5], Lloyd (1975) [3], and Waldman (1993) [2]. Although Kopeika (1998) [6] does not focus on infrared sensors, he does cover many of the topics treated in this book. Mathematics useful for electro-optics is given in Andrews (2003) [7] and Bracewell (2000) [8]. Linear systems are discussed by Lathi (2005) [9], Strum (2000) [10], and Gaskill (1978) [11], but only the book by Gaskill deals with linear systems in the context of optics. There are many books on optics. Books we have found useful include Hecht (2001) [1], Jenkins (2001) [12], and Pedrotti (2006) [13]. One of our favorite books for detectors is Dereniak (1996) [14]. Burle (1974) [15] gives a succinct treatment of radiometry and photometry. A more complete treatment of radiometry and photometry is given by McCluney (1994) [16]. The material in this text on testing and evaluation is supplemented by that in Holst (2008) [17]. Encylopedic treatments of material presented in this book are given by Driggers (2003) [17], Wolfe (1993) [19], and Zissis (1993) [20].

References

[1] Hecht, E., *Optics*, Boston, MA: Addison-Wesley, 2001.

[2] Waldman, G., and J. Wooten, *Electro-Optical Systems Performance Modeling*, Norwood, MA: Artech House, 1993.

[3] Lloyd, J., *Thermal Imaging Systems,* New York City, NY: Plenum Press, 1975.

[4] Daniels, A., *Field Guide to Infrared Systems, Detectors and FPAs*, SPIE, 2010.

[5] Holst, G. C., *Electro-Optical Imaging System Performance*, Bellingham, WA: SPIE, 2008.

[6] Kopeika, N. S., *A System Engineering Approach to Imaging*, Bellingham, WA: SPIE, 1998.

[7] Andrews, L. C., and R. L. Phillips, *Mathematical Techniques for Engineers and Scientists*, Bellingham, WA: SPIE, 2003.

[8] Bracewell, R. N., *The Fourier Transform and Its Applications*, Columbus, OH: McGraw-Hill, 2000.

[9] Lathi, B. P., *Linear Systems and Signals*, New York City, NY: Oxford University Press, 2005.

[10] Strum, R. D., and D. E. Kirk, *Contemporary Linear Systems Using MATLAB*, Salt Lake City, UT: Brooks/Cole, 2000.

[11] Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics*, John Wiley & Sons, 1978.

[12] Jenkins, F. A., and H. E. White, *Fundamentals of Optics*, Columbus, OH: McGraw-Hill, 2001.

[13] Pedrotti, F. L., L. M. Pedrotti, and L. S. Pedrotti, *Introduction to Optics*, Englewood Cliffs, NJ: Prentice Hall, 2006.

[14] Dereniak, E. L., and G. D. Boreman, *Infrared Detectors and Systems*, John Wiley & Sons, 1996.

[15] Burle, *Electro-Optics Handbook*, 1974, www.burle.com/onlineorderem.htm.

[16] McCluney, R., *Introduction to Radiometry and Photometry*, Norwood, MA: Artech House, 1994.

[17] Holst, G. C., *Testing and Evaluation of Infrared Systems*, Bellingham, WA: SPIE, 2008.

[18] Driggers, R. G., *Encyclopedia of Optical Engineering*, Monticello, NY: Dekker, 2003.

[19] Wolfe, W., and G. Zissis, *The Infrared Handbook,* Environmental Research Institute of Michigan, Ann Arbor, MI, 1993.

[20] Zissis, G. J., *The Infrared and Electro-Optical Systems Handbook*, Environmental Research Institute of Michigan, Ann Arbor, MI, 1993.

CHAPTER 2

Mathematics

This chapter is designed to provide the reader with the mathematical tools required for the modeling and analysis of infrared and EO sensor systems. The development of these models will allow the practitioner to predict system performance, arguably the single most important ingredient in sensor design. It is assumed the reader has an undergraduate background in engineering or physics and has some familiarity with Fourier analysis and linear system theory where convolution and correlation operators play a prominent role. Linear operators can be used to capture much of the physics needed in describing components such as lenses, detectors, and displays. The mathematical background provided here will also prove useful in understanding and modeling diffraction effects (Chapter 4).

Much of the emphasis in this chapter is on transitioning the principles of linear system theory from one to two dimensions, as imaging systems tend to be modeled with 2D functions. The transition presented here is not a rigorous one; however, it will allow the sensor analyst to develop many useful models and perform various design analyses. More in-depth discussions on complex numbers, functions, and transforms applied to optical systems can be found in Goodman [11] and Gaskill [10]. References by Ross [20] and Stemler [22] on communications theory may also prove useful to the reader since the mathematics is similar. Goodman, for example, states that "optics has gradually developed ever stronger ties with communications and information sciences." Other good references on the mathematical modeling of optical systems are given at the end of the chapter. We begin this chapter with a review of complex functions.

2.1 Complex Functions

The sensing systems described in this book are all designed to measure the electromagnetic field generated by various sources. The propagation of this field through air (or other media) is often described mathematically by a sinusoid (or combination of sinusoids) with temporally and/or spatially varying phase. This description naturally lends itself to the use of complex functions, or phasors [12]. A complex function can be represented in the form

$$g(x) = v(x) + jw(x) \tag{2.1}$$

or in the form

$$g(x) = a(x)e^{j\Phi(x)} \tag{2.2}$$

Equation (2.1) gives the complex function in Cartesian coordinates where the real part of the function is $v(x) \equiv \text{Re}\{g(x)\}$ and $w(x) \equiv \text{Im}\{g(x)\}$ is the imaginary part of the function. Equation (2.2) gives the function in polar form, where the non-negative amplitude of the function is $a(x)$ and $\varnothing(x)$ is the argument or phase of the function. The complex conjugate of the function is denoted $g^*(x)$ and is obtained by changing the sign on the imaginary part of (2.1) or the phase of (2.2). The relationships between the Cartesian and the polar forms of the function are summarized in Table 2.1.

The polar-to-Cartesian coordinate conversion is determined using Euler's relationship [21]

$$e^{j\phi(x)} = cos(\phi(x)) + jsin(\phi(x)) \tag{2.3a}$$

where $cos(\phi(x))$ is the real part of the function and $sin(\phi(x))$ is the imaginary part of the function. Euler's relationship can be easily manipulated to also write

$$sin(\Phi(x)) = \frac{e^{j\Phi(x)} - e^{-j\Phi(x)}}{2j} \tag{2.3b}$$

and

$$cos(\Phi(x)) = \frac{e^{j\Phi(x)} + e^{-j\Phi(x)}}{2} \tag{2.3c}$$

In a particular model we may be interested in how the waveform (modeled as a complex function) is altered by the atmosphere as it travels toward a detector. Modeling these phenomena requires the practitioner to be familiar with the manipulation of complex functions. The operations that will be encountered here consist of addition, subtraction, multiplication, division, and convolutions. Addition and subtraction of complex functions are generally accomplished in Cartesian coordinates with both the real and imaginary parts adding or subtracting directly. Multiplication and division are more tractable in polar coordinates, where

$$\frac{a_1(x)e^{-j\Phi_1(x)}}{a_2(x)e^{-j\Phi_2(x)}} = \frac{a_1(x)}{a_2(x)}e^{-j[\Phi_1(x)-\Phi_2(x)]} \tag{2.4}$$

$$a_1(x)e^{-j\Phi_1(x)}a_2(x)e^{-j\Phi_2(x)} = a_1(x)a_2(x)e^{-j[\Phi_1(x)+\Phi_2(x)]} \tag{2.5}$$

Note that while vector division is not defined, complex function division is defined providing $a_2(x) \neq 0$. Transform and convolution operations are addressed later in this chapter.

Table 2.1 Cartesian and Polar Relationships

$a(x) = \lvert g(x)\rvert = \sqrt{v^2(x) + w^2(x)}$
$\phi(x) = arg\left[g(x)\right] = tan^{-1}\left[\frac{w(x)}{v(x)}\right]$
$v(x) = \text{Re}\{g(x)\} = a(x)\ cos(\phi(x))$
$w(x) = \text{Im}\{g(x)\} = a(x)\ sin(\phi(x))$

2.2 Common One-Dimensional Functions

A number of one-dimensional (1D) functions must be understood before functions can be extended into two dimensions. The discussion here has been limited to those that are useful in infrared and EO sensor analysis. Table 2.2 lists these functions along with their definitions. These functions are frequently used to model the behavior of system components such as apertures, detectors, and analog to-digital samplers. For example, the *Gaus*(x) function may depict the spatial distribution of the intensity in a cathode-ray tube (CRT) display spot. More often, however, the analyst is required to scale, dilate, or translate the listed functions in order to successfully model a specific component.

Consider, for example, the function shown in Figure 2.1. Compare it with the definition of the *rect* function given in Table 2.2.

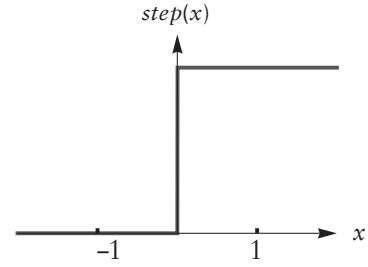
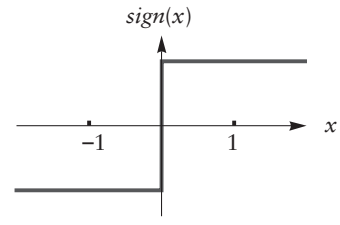
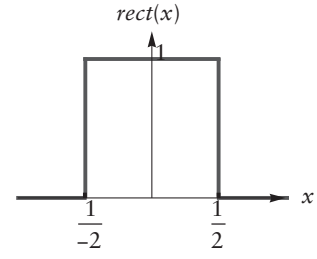
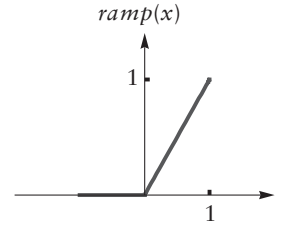
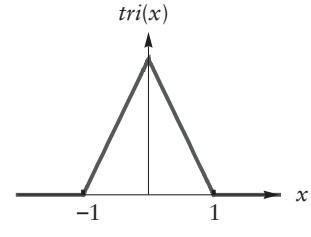
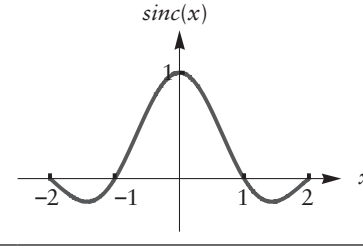
The amplitude of the function has been scaled by two and the origin translated to $x = 1$. Finally, the width of the function has been dilated by a factor of 4 so that the width of the rectangle is four units. Note that when x is set to 3, the argument of the *rect* function becomes 1/2, a discontinuous transition point for the *rect* function. Because integrating over a specific point does not contribute to an integral (unless that point is a delta function), the value of a function at a discontinuity can be chosen freely. A common choice is to pick a value that is the mean of the function just to the left and just to the right of the discontinuity. In general, a function $f(x)$ can be scaled, translated, and dilated via $f(x) \rightarrow af\left(\frac{x - x_0}{b}\right)$ where a is the scaling constant, b controls dilation (width), and the origin is shifted to x_0.

There is no substitute for practice in the manipulation of functions, so a number of problems are given at the end of this chapter. For example, it should become obvious that the *rect*(x) function can be represented as *rect*(x) = *step*(x + 1/2) *step*(1/2 – x) or as *rect*(x) = *step*(x + 1/2) – *step*(x – 1/2). Another worthwhile exercise is to represent the *tri*(x) function with a number of *ramp*(x) functions. The end goal, of course, is to choose the simplest possible representation for the particular problem at hand. Representation in one form may cause a great deal of unnecessary mathematics (e.g., when transformations are required), whereas representation in a different way may provide a straightforward evaluation.

The delta function, sometimes called the impulse or the Dirac function, is one of great practical importance in describing the performance of infrared and EO system components. It is used widely by both engineers and mathematicians in a number of fields. It is not a true function because it is not defined when its argument is zero. See Hsu [14] for a discussion of generalized functions and their relationship to the delta function. The impulse function is characterized by infinite height, zero width, and possesses an area of unity. In fact, the definition given in Table 2.2 could be modified with the use of *rect* or *tri* functions instead of the *Gaus* function. The representation is designed to have vanishing width but area equal to one. Regardless of how the delta function is represented, it has three very important properties:

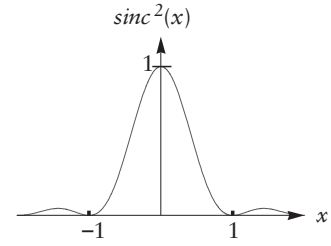
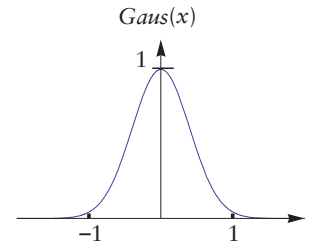
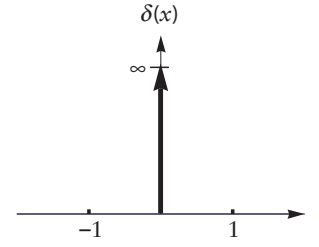
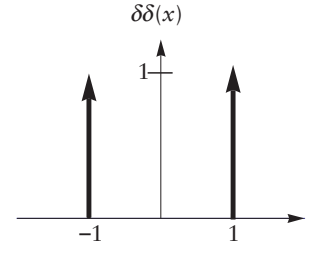
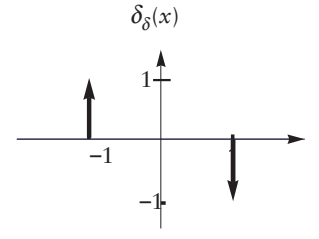
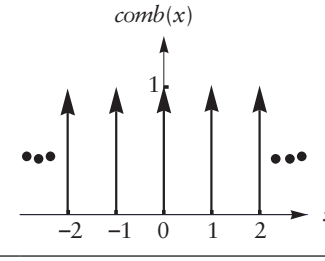
$$\delta(x - x_0) = 0 \quad x \neq x_0 \tag{2.6a}$$

Table 2.2 One-Dimensional Functions

Name	*Function*	*Definition*	*Graph*
Step	*step* (x)	$0,\ x<0$ $\frac{1}{2},\ x=0$ $1,\ x>0$	
Signum	*sign* (x)	$1, x>0$ $0, x=0$ $-1, x<0$	
Rectangle	*rect* (x)	$1, \lvert x\rvert \le \frac{1}{2}$ $0, \lvert x\rvert > \frac{1}{2}$	
Ramp	*ramp* (x)	$0, x \le 0$ $x, x>0$	
Triangle	*tri* (x)	$1-\lvert x\rvert, \lvert x\le 1\rvert$ $0, \lvert x\rvert > 1$	
Sinc	*sinc* (x)	$\frac{\sin(\pi x)}{\pi x}$	

(*continued*)

Table 2.2 Continued

Name	*Function*	*Definition*	*Graph*
sinc^{2}	*sinc*^{2} (x)	$\left(\frac{sin(\pi x)}{\pi x}\right)^2$	
Gaussian	*Gaus* (x)	$e^{-\pi x2}$	
Delta	$\delta\,(x)$	$lim_{b\to 0}\frac{1}{\lvert b\rvert}Gaus\left(\frac{x}{b}\right)$	
Even impulse pair	$\delta\delta\,(x)$	$\delta\,(x+1)+\delta(x-1)$	
Odd impulse pair	$\delta_\delta\,(x)$	$\delta\,(x+1)-\delta(x-1)$	
Comb	*Comb* (x)	$\sum_{n=-\infty}^{\infty}\delta(x-n)$	

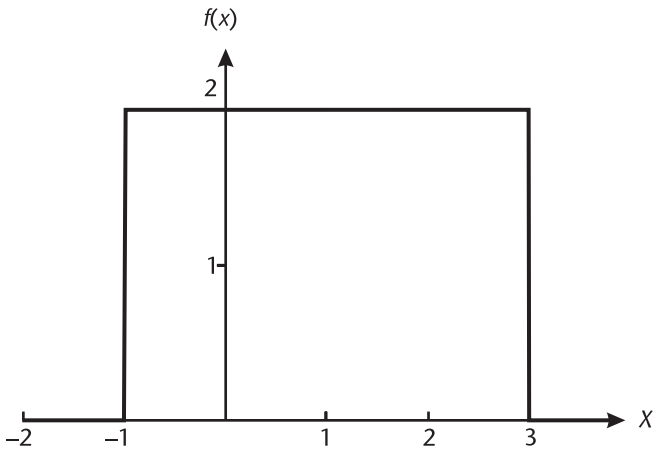


Figure 2.1 $f(x) = 2\,rect\left(\frac{x-1}{4}\right)$.

$$\int_{x_1}^{x_2} g(\alpha)\delta(\alpha - x_0)d\alpha = g(x_0),\ x_1 < x_0 < x_2 \tag{2.6b}$$

$$\delta\left(\frac{x - x_0}{b}\right) = |b|\delta(x - x_0) \tag{2.6c}$$

Three related functions are composed of delta functions. The first two are the even impulse pair and the odd impulse pair as shown in Table 2.2. These functions are simply the summation of two impulse functions, where the first is an even function configuration and the second is an odd function configuration. Delta functions have a number of uses in engineering practice. The Fourier transform of a harmonic signal with fixed frequency is given by a delta function placed at the frequency of oscillation. Additionally, the *comb* function is an infinite sum of delta functions that can represent the sampling nature of discrete detectors in space or a digital sampling process in the time domain.

2.3 Two-Dimensional Functions

Two-dimensional functions play a key role in the modeling and evaluation of infrared and EO sensors. Most infrared and EO systems form an image at a 2D plane (e.g., a detector array or film). The complex wave propagating through the atmosphere and the sensor can be represented in a 2D (transverse to the direction of propagation) plane as well. As a result, sensor components can usually be described by 2D functions that have some symmetry in either Cartesian or polar coordinates. Symmetry reduces the complexity of both the problem formulation and the mathematics necessary to evaluate the expression and should be invoked whenever possible. A special class of 2D functions is those that are separable in some particular coordinate system. A 2D function is separable in Cartesian coordinates if

$$g(x,y) = g_X(x)g_Y(y) \tag{2.7}$$

and in polar coordinates ($r = \sqrt{x^2 + y^2}, \theta = (tan^{-1}(\frac{y}{x}))$ if

$$g(r,\theta) = g_R(r)g_\Theta(\theta) \tag{2.8}$$

In simple terms, a separable function is one that can be written as the product of 1D functions of orthogonal independent variables. Separable functions make the analysis of 2D systems much simpler. For example, the 2D Fourier transform of a separable function is the product of the 1D Fourier transforms. An example of a separable 2D function is shown in Figure 2.2. The top graph shows the function,

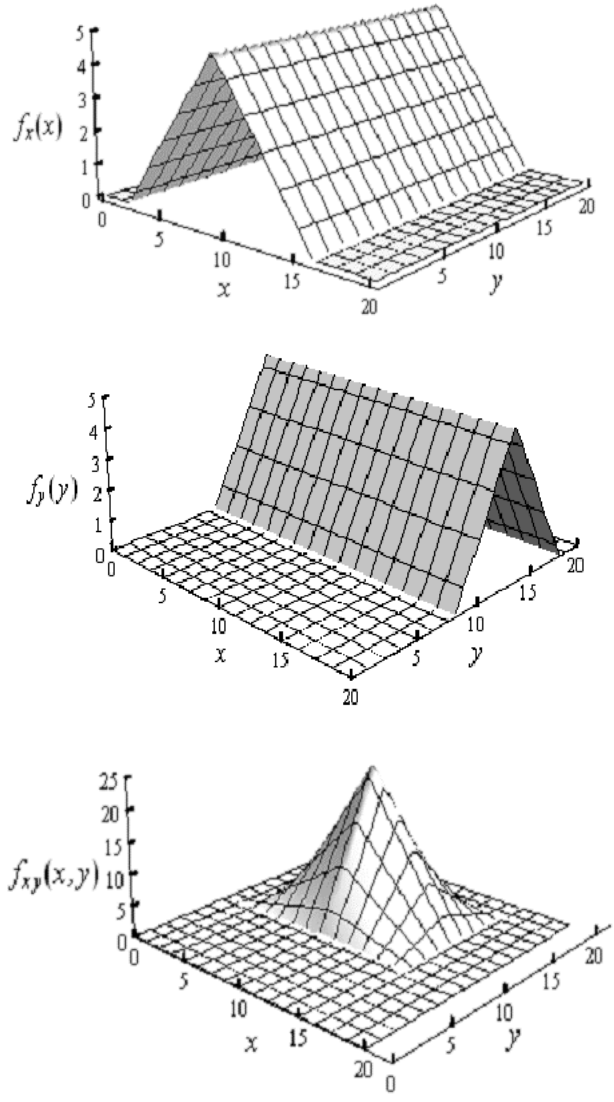


Figure 2.2 Two-dimensional triangle function.

$f_X(x) = 5\,tri\left(\frac{x-9}{7}\right)$ and the second graph shows $f_Y(y) = 5\,tri\left(\frac{y-13}{5}\right)$. The two functions multiplied give $25\,tri\left(\frac{x-9}{7}, \frac{x-13}{5}\right)$, as shown in the bottom graph. Be careful about making generalizations about separable function shapes. The product of two triangle functions may, at first thought, be conceived as a pyramid. The product of the two triangle functions' values at every point along the x and y axes causes the unexpected curved shape along the diagonal. Only when viewed as slices parallel to the coordinate axes is the true shape of a triangle present because the multiplier of the triangles along each line is a constant.

As one can imagine, an infinite number of separable functions exists. Just using the functions shown in Table 2.2, we would have 144 different combinations of functions where a function in the x direction is multiplied by a function in the y direction. In fact, sums and differences of functions in one direction multiplied by sums and differences of functions in the other direction give 2D functions that are separable. Shifting and scaling of functions provide even more combinations of separable functions.

Three 2D functions commonly used in image modeling are the *cylinder*, *Gaussian*, and *sombrero* functions or $cyl(r)$, $Gaus(r)$, and $somb(r)$. Despite the single argument, these functions are two dimensional because $f(r) = f_r(\sqrt{x^2 + y^2})$, that is, $f_\theta(\theta) = 1$. That is, they are circularly symmetric, where r is the radial variable. The three functions are shown and defined in Figure 2.3. The $cyl(r)$ function is useful in representing circular aperture stops, pupils, and mirrors. The $Gaus(r)$ function is useful in modeling laser spot intensities and electronic beam written display spots. Finally, the $somb(r)$ function is similar to the $sinc(x)$ function, but in two dimensions and has the useful attribute of being the Fourier transform of the $cyl(r)$ function. The function J_1 shown in the definition of the *sombrero* function is a first-order Bessel function of the first kind, as given by Andrews [1, 2].

There are instances when the required functions are not separable, for example

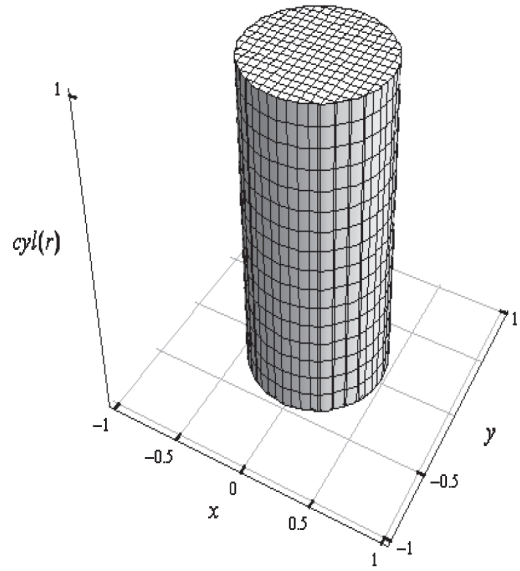
$$g(x,y) = rect\left(\frac{x-3}{3}, \frac{y-2}{2}\right) + 2\,rect\left(\frac{x+3}{4}, \frac{y-3}{3}\right)$$

While these functions may defy an analytical representation, if necessary a numerical representation can usually be found. Finally, we should note that a function that is separable in polar coordinates does not mean the function is separable in rectangular coordinates (they usually are not) and vice versa.

2.4 Convolution and Correlation

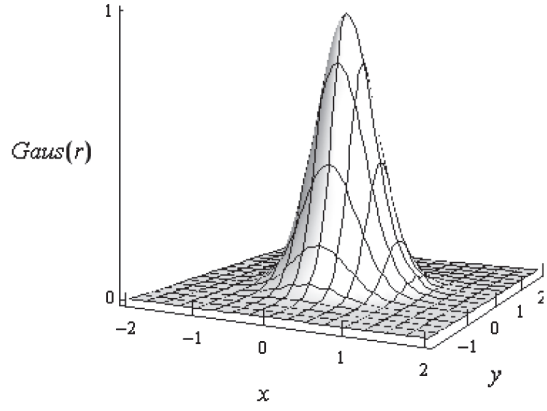
As stated in the introduction, this text will follow a linear systems approach to sensor modeling. (Chapter 3 describes this approach in some detail.) The chief assumption in this approach is that the output (response) of a given system component can be modeled as a linear combination of system inputs [see, e.g., (1.1) in Chapter 1].

Two operators that embody this property are the convolution and correlation operators. The convolution operation is defined mathematically as



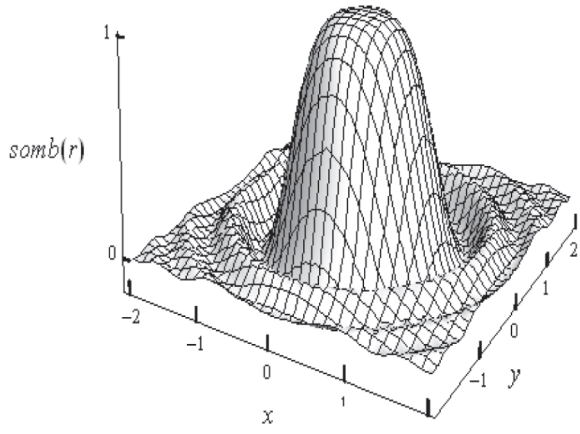
Cylinder function

$$cyl(r) = \begin{cases} 1 & r < 1/2 \\ 0 & r \geq 1/2 \end{cases}$$



Gaussian function

$$Gaus(r) = e^{-\pi r^2}$$



Sombrero function

$$somb(r) = 2J_1(\pi r)/(\pi r)$$

Figure 2.3 Circularly symmetric functions.

$$g(x) = f(x) * h(x) \equiv \int_{-\infty}^{\infty} f(\alpha)h(x - \alpha)\, d\alpha \tag{2.9}$$

In the context of linear systems theory this expression states that the response $g(x)$ to an input $f(x)$ acted on by a component modeled by $h(x)$ is an infinite superposition over all possible values of the function. Heuristically, the convolution of two functions $f(x)$ and $h(x)$ can be described by the process of "flip, slip, multiply, and integrate" [5]. Pick one function as stationary, usually the more complex of the two. Then, the flip is an inversion from left to right [around the $f(x)$ or $h(x)$ axis], the slip is the moving of one function along the x axis with respect to the other, the multiply is a point-by-point multiplication of the slipped function by the stationary function, and the integration is the summation of all multiplied points. This process is shown in Figure 2.4, where $f(x) = h(x) = rect(x)$.

The properties of the convolution integral include commutative:

$$f(x) * h(x) = h(x) * f(x) \tag{2.10}$$

distributive:

$$[af(x) + bh(x)] * g(x) = af(x) * g(x) + bh(x) * g(x) \tag{2.11}$$

associative:

$$[f(x) * g(x)] * h(x) = f(x) * [g(x) * h(x)] \tag{2.12}$$

and shift invariance:

$$f(x - x_0) * h(x) = g(x - x_0) \tag{2.13}$$

Another useful relationship is obtained when convolving an arbitrary function $f(x)$ with the delta function

$$f(x) * \delta(x) = f(x) \tag{2.14}$$

Convolution can be thought of as a filtering operation whereby the input $f(x)$ is "filtered" by the system $h(x)$. As a result, the output of a convolution tends to be smooth compared to the original functions (with the exception of convolutions involving delta functions). Convolution is useful in representing many physical phenomena, including the scanning of a rectangular-shaped detector across an inverted, or flipped, image. The convolution of complex functions is found in the same manner as that of (2.9), where the equation for the complex polar form is

$$\begin{aligned} f_3(x) = f_1(x) * f_2(x) &= \int_{-\infty}^{\infty} \alpha_1(\alpha)e^{j\varnothing_1(\alpha)}\alpha_2(x - \alpha)e^{j\varnothing_2(x-\alpha)}\, d\alpha \\ &= \int_{-\infty}^{\infty} \alpha_1(\alpha)\alpha_2(x - \alpha)e^{j[\varnothing_1(\alpha)+\varnothing_2(x-\alpha)]}\, d\alpha \end{aligned} \tag{2.15}$$

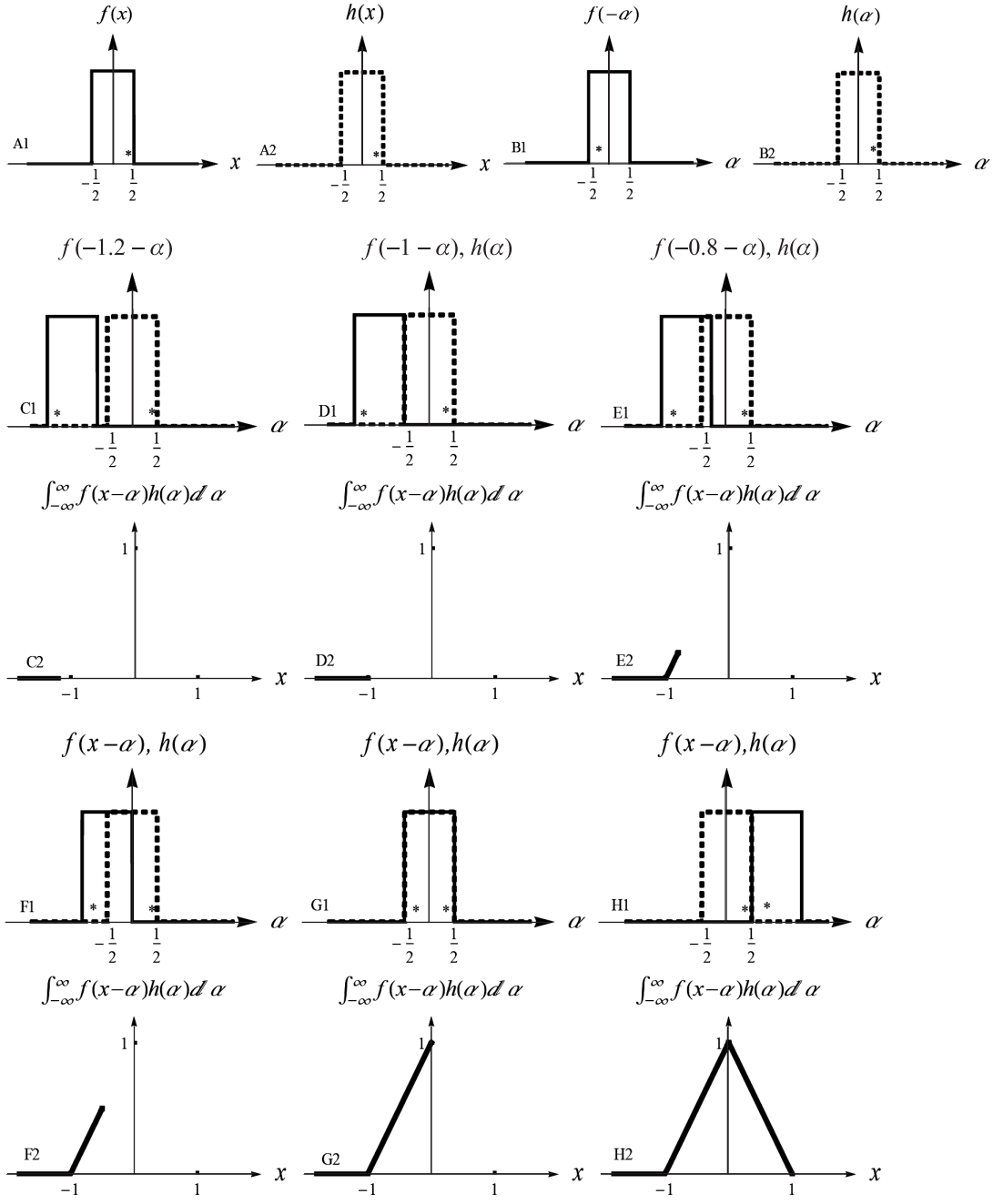


Figure 2.4 The process of convolution. A: Original *f* and *h* functions are *rect*(*x*). B: Flip *f* function and leave *h* function unchanged. C – H: Slip, multiply, and integrate with *x* successively taking the values –1.2, –1.0, –0.8, –0.5, 0, and 1.

In the complex Cartesian form

$$\begin{aligned} f_3(x) &= f_1(x) * f_2(x) = [v_1(x) + jw_1(x)] * [v_2(x) + jw_2(x)] \\ &= [v_1(x) * v_2(x) - w_1(x) * w_2(x)] + j[v_1(x) * w_2(x) + w_1(x) * v_2(x)] \\ &= v_3(x) + jw_3(x) \end{aligned} \tag{2.16}$$

Two-dimensional convolution in Cartesian coordinates is given by

$$g(x,y) = f(x,y) ** h(x,y) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha,\beta) h(x-\alpha, y-\beta)\, d\alpha\, d\beta \tag{2.17}$$

A nice convolution property of separable 2D functions is

$$g(x,y) = [f_X(x) f_Y(y)] ** [h_X(x) h_Y(y)] = [f_X(x) * h_X(x)][f_Y(y) * h_Y(y)] \tag{2.18}$$

The convolution of 2D functions in polar coordinates is limited here to those functions that are circularly symmetric. The convolution integral for two circularly symmetric functions in polar coordinates is

$$g_R(r) = f_R(r) ** h_R(r) = \int_0^{\infty} \int_0^{2\pi} f_R(r') h_R\left(\sqrt{r^2 + r'^2 - 2rr' \cos\theta'}\right) r'\, d\theta' dr' \tag{2.19}$$

A close cousin of the convolution is the *correlation* operator, defined as

$$g(x) = f(x) \star h(x) \equiv \int_{-\infty}^{\infty} f(\alpha) h(\alpha - x) d\alpha \tag{2.20a}$$

This operator is identical to convolution but with the argument of $h(x)$ reversed; that is, the "flip" operation is eliminated from the process. The two operators are clearly related by

$$g(x) = f(x) \star h(x) = f(x) * h(-x) \tag{2.20b}$$

The result of a correlation operation can be thought of as a measure of how similar one signal or function is to another. Due to the symmetry of the functions in Figure 2.4 (i.e., for even functions), the convolution is the same as correlation. Cross-correlation is also defined for complex valued functions as

$$\gamma_{fh}(x) \equiv f(x) \star h^*(x) \tag{2.21}$$

where $h^*(x)$ denotes the complex conjugate of $h(x)$. Finally, the complex autocorrelation is defined as

$$\gamma_f(x) \equiv f(x) \star f^*(x) \tag{2.22}$$

Correlation taken in two dimensions is a straightforward generalization of one dimensional correlation:

$$g(x,y) = f(x,y) \star\star h(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)h(\alpha - x, \beta - y)d\alpha d\beta$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha + x, \beta + y)h(\alpha,\beta)d\alpha d\beta \tag{2.23}$$

The 2D cross-correlation is not commutative but can be expressed as

$$f(x,y)\star\star h(x,y) = f(x,y) * * h(-x,-y) \tag{2.24}$$

The complex cross-correlation is defined as

$$\gamma_{fh}(x) \equiv f(x,y)\star\star h^*(x,y) \tag{2.25}$$

The complex autocorrelation can also be calculated in two dimensions:

$$\gamma_f(x) \equiv f(x,y)\star\star f^*(x,y) \tag{2.26}$$

In 2D polar coordinates, we again limit the correlations to circularly symmetric functions with the result that the correlation reduces to a convolution. That is,

$$f_R(r)\star\star f_R(r) = f_R(r) * * f_R(r) \tag{2.27}$$

With the lack of a flip operator, the correlation is used to model actions such as a noninverted image that is scanned by a detector. The complex autocorrelation is used in operations such as the determination of the optical transfer function of a sensor. Both the convolution and the correlation functions provide measures of likeness between two functions, where the convolution likeness involves one function that is flipped.

2.5 The Fourier Transform

A mathematical tool of great importance in the analysis of IR and EO systems is the Fourier transform [10]. This particular tool describes the frequency or spectral content of spatial variations in the image and can describe the throughput characteristics of the IR and EO systems. An object comprised of many small features that are closely spaced with good contrast has some high spatial frequency content, whereas large objects that are spaced far apart have significant low spatial frequency content. The Fourier transform of an object space or an image space quantifies their corresponding small and large spatial variations. Fourier transforms are useful in modeling sensors and describing the spatial frequency content of an image that traverses each of the sensor components. Sensor optics, detectors, electronics, and displays can be represented by the Fourier transform of their spatial models. Exceptions are those electronics that cause signals to increase or decrease exponentially; these must be represented with the Laplace transform. Many electronics, however,

can be adequately described by use of the Fourier transform. The Fourier transform of a function $f(x)$ that can be real or complex is

$$F(\xi) = \int_{-\infty}^{\infty} f(x)e^{-j2\pi\xi x}dx \tag{2.28}$$

where the existence of the Fourier transform is met by the requirement that it be applied to physically realizable systems. The spatial frequency is given in cycles per milliradian or cycles per meter depending on the units of x. Gaskill [10] states that a few functions that are generally used by engineers, such as the impulse and infinite periodic functions, are not physically realizable; however they can still be useful in the analysis of systems because they are good approximations. Many of these functions such as the delta function do have generalized Fourier transforms that will be used throughout this book. Given a Fourier transform, the inverse Fourier transform can be determined by

$$f(x) = \int_{-\infty}^{\infty} F(\xi)e^{j2\pi x\xi}d\xi \tag{2.29}$$

where this function definition completes what is called the Fourier transform pair. Note that the only difference between these two transforms is the sign of the exponential kernel within the integrals. To avoid writing the above integrals each time we want to express the Fourier transform operation, a standard shorthand for the Fourier transform operation is used:

$$F(\xi) = \mathcal{F}\{f(x)\} \text{ and } f(x) = \mathcal{F}^{-1}\{F(\xi)\} \tag{2.30}$$

The first operation is the Fourier transform, and the second operation is the inverse Fourier transform. Often the original function is lowercase and the transformed function is uppercase.

The extension of Fourier transforms to two dimensions is a simple one in Cartesian coordinates. The forward Fourier transform is

$$\mathcal{F}\mathcal{F}\{f(x,y)\} = F(\xi,\eta) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi[\xi x+\eta y]}\,dx\,dy \tag{2.31}$$

and the inverse Fourier transform is

$$\mathcal{F}\mathcal{F}^{-1}\{F(\xi,\eta)\} = f(x,y) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi,\eta)e^{j2\pi[x\xi+y\eta]}\,d\xi\,d\eta \tag{2.32}$$

Separable functions, again, have a nice property regarding the forward and reverse Fourier transforms

$$\mathcal{F}\mathcal{F}\{f(x,y)\} = \mathcal{F}\{f_X(x)\}\mathcal{F}\{f_Y(y)\} \tag{2.33}$$

so that the individual 1D transforms can be calculated separately and then multiplied. The inverse transforms of separable functions can be determined in the same manner. In 2D polar coordinates, the equivalent Fourier transform is called the Fourier-Bessel transform, or the Hankel transform of zero order [12, 23]:

$$G(\rho) = 2\pi \int_0^\infty r\, g_R(r) J_0(2\pi\rho r)\, dr \tag{2.34}$$

The function $g_R(r)$ is a 2D polar function that is circularly symmetric [recall that $g_\theta(\theta) = 1$]. The J_0 term is the zero-order Bessel function of the first kind. The inverse transform can be found by

$$g_R(r) = 2\pi \int_0^\infty \rho G(\rho) J_0(2\pi r\rho)\, d\rho \tag{2.35}$$

Note that the inverse transform is identical to the forward transform. The shorthand notations for the Fourier-Bessel transform and corresponding inverse transform are $\mathcal{BB}\{g_R(r)\}$ and $\mathcal{BB}^{-1}\{G(\rho)\}$.

2.6 Properties of the Fourier Transform

A number of properties are useful in the evaluation and manipulation of Fourier transforms. These properties are not just for academic interest, but are used widely in the analysis of IR and EO systems. The mathematics are shown as the following theorems and, where applicable, the theorems assume that $\mathcal{FF}\{f(x, y)\} = F(\xi, \eta)$ and $\mathcal{FF}\{h(x, y)\} = H(\xi, \eta)$.

Linearity theorem
The transform of the sum of two functions is the sum of their individual transforms:

$$\mathcal{FF}\{af(x,y) + bh(x,y)\} = a\mathcal{FF}\{f(x,y)\} + b\mathcal{FF}\{h(x,y)\} = aF(\xi,\eta) + bH(\xi,\eta) \tag{2.36}$$

where a and b are constants.

Similarity theorem
"Wide" functions in the space domain correspond to "narrow" functions in the frequency domain and vice versa. This width characteristic is coupled to a change in amplitude to conserve energy (i.e., see Parseval's theorem below):

$$\mathcal{FF}\{f(ax,by) = \frac{1}{|ab|} F\left(\frac{\xi}{a}, \frac{\eta}{b}\right) \tag{2.37}$$

Shift theorem
The shifting of a function in space corresponds to a phase shift in the frequency domain:

$$\mathcal{FF}\{f(x-a,y-b)\} = F(\xi,\eta)e^{-j2\pi(a\xi+b\eta)} \tag{2.38}$$

Parseval's theorem
Energy is conserved in both the space and frequency domains:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x,y)|^2\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F(\xi,\eta)|^2\, d\xi\, d\eta \tag{2.39}$$

Convolution theorem
The convolution of two functions in the space domain corresponds to the multiplication of the transforms in the frequency domain:

$$\begin{aligned}\mathcal{FF}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha,y-\beta)\,d\alpha\,d\beta\right\} &= \mathcal{FF}\, f(x,y) * h(x,y)\\ &= F(\xi,\eta)H(\xi,\eta)\end{aligned} \tag{2.40}$$

Correlation theorem
The transform of the complex autocorrelation is the square of the transform's magnitude:

$$\mathcal{FF}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)f^*(\alpha-x,\beta-y)\,d\alpha\,d\beta\right\} = \mathcal{FF}\{\gamma_f(x,y)\} = |F(\xi,\eta)|^2 \tag{2.41}$$

Fourier integral theorem
The transform of an inverse transform and the inverse transform of a transform yield the original function:

$$\mathcal{FF}\left\{\mathcal{FF}^{-1}\{f(x,y)\}\right\} = \mathcal{FF}^{-1}\left\{\mathcal{FF}\{f(x,y)\}\right\} = f(x,y) \tag{2.42}$$

The properties described here are very basic, and each property has a duality characteristic. That is, each one of the properties can be applied to not only the forward transform, but also to the inverse transform. These properties are also applicable to the polar 2D functions, where the similarity theorem is modified to conserve energy and the shifting theorem is held in Cartesian coordinates. Table 2.3 lists a summary of the Fourier transform properties.

2.7 Transform Pairs

Each of the functions described in Sections 2.2 and 2.3 has a corresponding Fourier transform. The function and its corresponding Fourier transform are called a transform pair. Take, for example, the *rect*(x) function of Table 2.2. The Fourier transform can be applied to the function to yield

Table 2.3 Fourier transform properties

Space Domain	*Frequency Domain*
$f(\pm x, \pm y)$	$F(\pm\xi, \pm\eta)$
$f^*(\pm x, \pm y)$	$F^*(\pm\xi, \pm\eta)$
$f_X(x)\, f_Y(y)$	$F_X(\xi)\, F_Y(\eta)$
$f\left(\frac{x}{a}, \frac{y}{b}\right)$	$\lvert ab\rvert\, F(a\xi, b\eta)$
$f(x \pm x_0, y \pm y_0)$	$F(\xi, \eta)\, e^{\pm j2\pi x_0 \xi}\, e^{\pm j2\pi y_0 \eta}$
$f(x, y) e^{\pm j2\pi\xi_0 x}\, e^{\pm j2\pi\eta_0 y}$	$F(\xi \pm \xi_0, \eta \pm \eta_0)$
$af(x, y) + bh(x, y)$	$aF(\xi, \eta) + bH(\xi, \eta)$
$f(x, y)$ ** $h(x, y)$	$F(\xi, \eta)\, H(\xi, \eta)$
$f(x, y)\, h(x, y)$	$F(\xi, \eta)$ ** $H(\xi, \eta)$
$f(x,y)$ ★★ $h(x, y)$	$F(\xi, \eta)\, H(-\xi, -\eta)$
$f(x, y)\, h(-x, -y)$	$F(\xi, \eta)$ ★★ $H(\xi, \eta)$
$\gamma_{fh} = f(x,y)$ ★★ $h^*(x, y)$	$F(\xi, \eta)\, H^*(\xi, \eta)$
$\gamma_f(x, y)\, h^*(x, y)$	$\gamma_{FH} = F(\xi,\eta)$ ★★ $H^*(\xi, \eta)$
$f(x, y)$ ★★ $f^*(x, y)$	$\lvert F(\xi, \eta)\rvert^2$
$\lvert f(x, y)\rvert^2$	$\gamma_f = F(\xi, \eta)$ ★★ $f^*(\xi, \eta)$
$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)\, d\alpha\, d\beta = F(0,0)$	$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi,\eta)\, d\xi\, d\eta = f(0,0)$
$g_R(r)$	$G(\rho)$
$g_R\left(\frac{r}{a}\right)$	$(\lvert a\rvert)^2\, G(a\rho)$
$ag_1(r) + bg_2(r)$	$aG_1(\rho) + bG_2(\rho)$
$g_1(r)$ ** $g_2(r) = g_1(\rho)$ ★★ $g_2(\rho)$	$G_1(\rho)\, G_2(\rho)$
$g_1(r)\, g_2(r)$	$G_1(\rho)$ ** $G_2(\rho) = G_1(\rho)$ ★★ $G_2(\rho)$
$\gamma_{g_1g_2}(r) = g_1(r)$ ★★ $g_2^*(r)$	$G_1(\rho)\, G_2^*(\rho)$
$g_1(r)g_2^*(r)$	$\gamma_{G_1G_2} = G_1(\rho)$★★$G_2^*(\rho)$
$\gamma_g(r) = g(r)$ ★★ $g^*(r)$	$\lvert G(\rho)\rvert^2$
$\lvert g_R(r)\rvert^2$	$\gamma_G(\rho) = G(\rho)$ ★★ $G^*(\rho)$
$g_R\left(\sqrt{x^2 + y^2}\right)$	$G\left(\sqrt{\xi^2 + \eta^2}\right)$
$\int_0^{\infty} g_R(r) r dr = G(0)$	$\int_0^{\infty} G(\rho)\rho d\rho = g_R(0)$

$$F(\xi) = \int_{-\infty}^{\infty} rect(x) e^{-j\,2\pi\,\xi x} dx = \int_{-1/2}^{1/2} e^{-j\,2\pi\,\xi x} dx$$

$$= \frac{1}{j\,2\pi\,\xi}\left[e^{j\,2\pi\frac{\xi}{2}} - e^{-j\,2\pi\frac{\xi}{2}}\right] \tag{2.43}$$

From Euler's equation of (2.3b), the transform becomes

$$F(\xi) = \frac{sin(\pi\,\xi)}{\pi\,\xi} = sinc(\xi) \tag{2.44}$$

This shows that the Fourier transform of $rect(x)$ is $sinc(\xi)$. In two dimensions, the transform pair of $rect(x, y)$ is $sinc(\xi, \eta)$.

Another important transform is that of the delta function. The transform can be found by the Fourier integral

$$F(\xi) = \int_{-\infty}^{\infty} \delta(x) e^{-j2\pi\xi x}\, dx \tag{2.45}$$

The function is equal to zero at locations other than $x = 0$, so the integral becomes

$$F(\xi) = \int_{-\infty}^{\infty} \delta(0) e^{-j2\pi\xi 0}\, dx = \int_{-\infty}^{\infty} \delta(0)\, dx = 1 \tag{2.46}$$

In a similar fashion the Fourier transform of $\delta(x, y)$ yields $F(\xi, \eta) = 1$.

Table 2.4 lists a number of common Fourier transform pairs for IR and EO application. The transform pair table along with the transform properties can often be used to model a function that adequately describes the EO and IR component or sensor of interest. Table 2.4 can be considered a lookup table where the transform pairs can be taken and used directly. The pairs are given only in the forward transform direction. However, the duality relationship

$$\mathcal{F}\{F(x)\} = f(-x) \tag{2.47}$$

doubles the size of the transform pair table

Table 2.4 Fourier transform pairs

Space domain	*Frequency domain*
1	$\delta(\xi)$
$\delta(x)$	1
$rect(x)$	$sinc(\xi)$
$\delta\delta(x) = [\delta(x + 1) + \delta(x - 1)]$	$2\, cos(2\pi\xi)$
$\delta_\delta(x) = [\delta(x + 1) - \delta(x - 1)]$	$2\, sin(2\pi\xi)$
$\delta(x, y) = \delta(x)\, \delta(y)$	1
$\delta(x \pm x_0, y \pm y_0)$	$e^{\pm j2\pi x_0 \xi} e^{\pm j2\pi y_0 \eta}$
$e^{\pm j2\pi\xi_0 x} e^{\pm j2\pi\eta_0 y}$	$\delta(\xi \mp \xi_0, \eta \mp \eta_0) = \delta(\xi \mp \xi_0)\delta(\eta \mp \eta_0)$
$cos(2\pi\xi_0 x)$	$\frac{1}{2}\left[\delta(\xi - \xi_0) + \delta(\xi + \xi_0)\right]$
$sin(2\pi\eta_0 y)$	$\frac{1}{2j}\left[\delta(\eta - \eta_0) - \delta(\eta + \eta_0)\right]$
$rect(x, y) = rect(x)\, rect(y)$	$sinc(\xi, \eta) = sinc(\xi)\, sinc(\eta)$
$tri(x, y) = tri(x)\, tri(y)$	$sinc^2(\xi, \eta) = sinc^2(\xi)\, sinc^2(\eta)$
$Gaus(x,y) = Gaus(x)\, Gaus(y)$	$Gaus(\xi, \eta) = Gaus(\xi)\, Gaus(\eta)$
$comb(x, y) = comb(x)\, comb(y)$	$comb(\xi, \eta) = comb(\xi)\, comb(\eta)$
$sgn(x, y) = sgn(x)\, sgn(y)$	$\frac{1}{j\pi\xi}\frac{1}{j\pi\eta}$
$cyl(r)$	$somb(\rho)$
$\frac{\delta(r)}{\pi r}$	1
$\delta(r - r_0)$	$2\pi r_0 J_0(2\pi r_0 \rho)$
$Gaus(r)$	$Gaus(\rho)$

2.8 Probability

Probability is helpful in many applications including searching and can be understood through a relative frequency of occurrence or through an axiomatic approach. In the relative frequency approach, if an event A can happen in n_A ways in n trials, then the probability P_A for the event is

$$P_A = lim_{n\to\infty} \frac{n_A}{n} \tag{2.48}$$

Of course n_A and n are positive so P_A is positive. In using Equation (2.48) it is assumed that all n events are equally likely.

It takes experience to identify the equally likely events. For example, consider an experiment in which two coins are tossed into the air and we wish to calculate the probability of observing one head and one tail after both coins have landed. One way of counting enumerates the possibilities this way: {two heads, two tails, one head, one tail}. If these outcomes were equally likely, then the probability of observing one head and one tail would be 1/3. Another way of counting identifies the coins and explicitly lists the outcomes for the first and second coin: {HH, TT, HT, TH}. If these possibilities were equally likely then the probability of observing one head and one tail in this experiment would be 2/4. Experiment shows that the events listed in the second method of counting are equally likely, so the probability of observing one head and one tail in the described experiment is 1/2 rather than 1/3. The hard part of probability theory is enumerating the outcomes in such a way that each outcome is equally likely. Before calculating a probability it is important to conceptually define the experiment and carefully decide which outcomes are going to be considered in the set A.

Equation (2.48) is useful for discrete events. For continuous events we define a *probability density function* $f(t)$ with the properties:

$$f(t) \geq 0 \tag{2.49}$$

$$\int_{-\infty}^{\infty} f(t)\, dt = 1 \tag{2.50}$$

$$P(t_0 < t < t_1) = \int_{t_0}^{t_1} f(t)\, dt \tag{2.51}$$

Equation (2.49) asserts that consistent with Eq. (2.48), $f(t)$ cannot be negative. Equation (2.51) asserts that for a continuous function the integral on the right-hand side equals the probability of observing the event in the interval between t_0 and t_1. Equation (2.50) asserts that all the possibilities must add up to one.

The symbol t suggests time and although this interpretation of the symbol is appropriate to search, the argument of the probability of the density function can be any physical quantity. For example, the Maxwell distribution gives the probability density function for the speed of air molecules.

An alternate description of probability is through the *cumulative distribution function* $F(t)$ defined by

$$F(t) \equiv \int_{-\infty}^{t} f(\alpha)\, d\alpha \tag{2.52}$$

The relationship between the cumulative distribution function and the probability density function is readily seen using a rule for differentiating integrals due to Leibnitz. Suppose

$$G(z) = \int_{a(z)}^{b(z)} g(x,z)\, dx \tag{2.53}$$

Then

$$\frac{dG(z)}{dz} = \frac{db(z)}{dz} g(b(z),z) - \frac{da(z)}{dz} g(a(z),z) + \int_{a(z)}^{b(z)} \frac{\partial g(x,z)}{\partial z}\, dx \tag{2.54}$$

Applying Leibnitz's rule to Equation (2.52) yields

$$\frac{dF(t)}{dt} = f(t) \tag{2.55}$$

Equation (2.52) shows how to go from the probability density function to the cumulative distribution function. Equation (2.55) shows how to go from the cumulative distribution function to the probability density function. Probability density functions and cumulative distribution functions are alternate approaches to doing probability calculations. Both approaches are useful.

A convention that is almost universally followed is to denote probability density functions with lowercase letters and to use uppercase letters to denote cumulative distribution functions. Frequently, probability density function and cumulative distribution functions are abbreviated pdf and CDF, respectively.

The concepts of *union* and *intersection* are useful for understanding probability theory. Sets A, B, C and D are defined by

$$A = \{a,b,c,d,e,f\}; \quad B = \{e,f,g,h,i,j\}; \quad C = \{g,h,i\}; \quad D = \{j,k\} \tag{2.56}$$

Then the intersection of sets A and B is

$$A \cap B = \{e,f\} \tag{2.57}$$

and the union of sets A and B is

$$A \cup B = \{a,b,c,d,e,f,g,h,i,j\} \tag{2.58}$$

The intersection of sets consists of those elements common to the two sets; the union of sets consists of those elements that are in either set. Although the union and intersection have been illustrated for sets consisting of discrete elements, the concepts are also valid for sets that have a continuum of elements. The *entire space*

S consists of every possible outcome. Assuming sets A, B, C, and D include every outcome, we get $S = \{a,b,c,d,e,f,g,h,i,j,k\}$

The probability of the entire space is one:

$$P(S) = 1 \tag{2.59}$$

Equation (2.50) illustrates Equation (2.59) for the continuum.

Two sets are said to be *mutually exclusive* if their intersection is empty; A and C are examples of mutually exclusive sets. For mutually exclusive sets the probability of the union is the sum of the individual probabilities:

$$P(A \cup C) = P(A) + P(C) \text{ assumes } A \cap C = \varnothing \tag{2.60a}$$

Equation (2.60a) says that for the mutually exclusive sets A and C the probability of A or C is the probability of A plus the probability of C. The sets A and B are not mutually exclusive. How would one calculate $P(A \cup B)$? The question is answered by (2.60b), a generalization of (2.60a):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.60b}$$

In (2.60b) note that in $P(A) + P(B)$ the elements $\{e, f\}$ are improperly counted twice; this is corrected by subtracting $P(A \cap B)$.

In contrast to the frequency approach of (2.48), the axiomatic approach theory uses (2.59), (260a), and the condition that probabilities cannot be negative to develop probability theory.

The concept of *conditional probability* is useful. The probability of A happening given that B has happened is given by:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.60c}$$

One reason conditional probabilities are useful is that the knowledge that B has happened influences the probability of A. Consider an example where $p(t)$ denotes the probability density of finding a target at time t where

$$p(t) = e^{-t} step(t), 0 < t < \infty \tag{2.61a}$$

One readily checks that $p(t)$ is a legitimate pdf since it satisfies (2.49) and (2.50). Let A be the probability of finding the target between 0 and 3 sec. Then using (2.51),

$$P(A) = \int_0^3 e^{-t}\, dt = 1 - e^{-3} \tag{2.61b}$$

Let B be the event that the target was not found between 0 and 1 sec. The event B corresponds to finding the target after $t = 1$:

$$P(B) = \int_1^\infty p(t)\, dt = e^{-1} \tag{2.61c}$$

The intersection of the events A and B corresponds to the time interval between 1 and 3:

$$P(A \cap B) = \int_1^3 p(t)\,dt = e^{-1} - e^{-3} \tag{2.61d}$$

Thus,

$$P(A|B) = \frac{e^{-1} - e^{-3}}{e^{-1}} = 1 - e^{-2} \tag{2.61e}$$

and we observe that $P(A) \neq P(A|B)$. Knowledge of B enabled a revised estimate for $P(A)$.

The concept of *independence* is very useful. Two events A and B are said to be independent if knowledge of event B does not influence the probability of A.

$$P(A|B) = P(A) \quad \text{if } A \text{ and } B \text{ are independent} \tag{2.62}$$

Consider an experiment where a coin is flipped and a die is tossed. The flipping of the coin does not influence the outcome of the die toss, so flipping a coin and tossing a die are independent events. For independent events A and B:

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)} \tag{2.63}$$

In (2.63) the first equality is valid because A and B are independent; the second equality is the definition of conditional probability. From (2.63) we conclude

$$P(A \cap B) = P(A)P(B) \quad \text{if } A \text{ and } B \text{ are independent} \tag{2.64}$$

The concepts of *random variable*, *mean*, *variance*, *standard deviation*, and *expected value* are ubiquitous in engineering and science. Let $f(t)$ denote the probability density for $\mathbf{t}$. When the outcome $\mathbf{t}$ of an experiment is uncertain and is conceptually described by a pdf or CDF, we call $\mathbf{t}$ a *random variable* and frequently denote it by bold type. Let $\mu_{\mathbf{t}}$ denote the mean of $\mathbf{t}$. Then

$$\mu_{\mathbf{t}} \equiv \int_{-\infty}^{\infty} t f(t)\,dt \tag{2.65a}$$

The concept of *expected value* provides another way to write (2.65a):

$$\mu_{\mathbf{t}} = E[\mathbf{t}] \tag{2.65b}$$

More generally, to calculate the expected value of a quantity in square brackets, multiply by the pdf associated with the random variable in square brackets and then integrate over the space of the pdf. The expected value of an arbitrary function of the random variable $\mathbf{t}$ is defined by

$$E[g(\mathbf{t})] = \int_{-\infty}^{\infty} g(t) f(t)\,dt \tag{2.65c}$$

Let $\sigma_{\mathbf{t}}^2$ denote the *variance* of **t**. Then

$$\sigma_{\mathbf{t}}^2 \equiv \int_{-\infty}^{\infty} (t - \mu_{\mathbf{t}})^2 f(t)\, dt \tag{2.66}$$

The positive square root of the variance is the *standard deviation*.

The *noise power spectral density* $S(f)$ is useful for defining the noise into a linear system and for calculating the noise out of a linear system. See Section 2.10 for more detailed information.

2.9 Important Examples

To illustrate the importance of the mathematical techniques described in this chapter, a few examples are given that are instrumental in sensor analysis.

Example 2.1

A linear detector array with 128 elements can be modeled in one dimension (the lengthwise dimension) to determine its voltage response to a particular image pattern. The detectors are uniform in response across their entire surface (say, 1V), square in shape, and 20 μm in height with 40-μm center-to-center spacing. The analyst must write a response function. A simple sketch is shown in Figure 2.5.

The function can be represented as a *rect* function convolved with a suitably constructed *comb* function. The useful format of the *comb* function is

$$\frac{1}{|a|} comb\left(\frac{x - x_0}{a}\right) = \sum_{n=-\infty}^{\infty} \delta(x - x_0 - n^a) \tag{2.67}$$

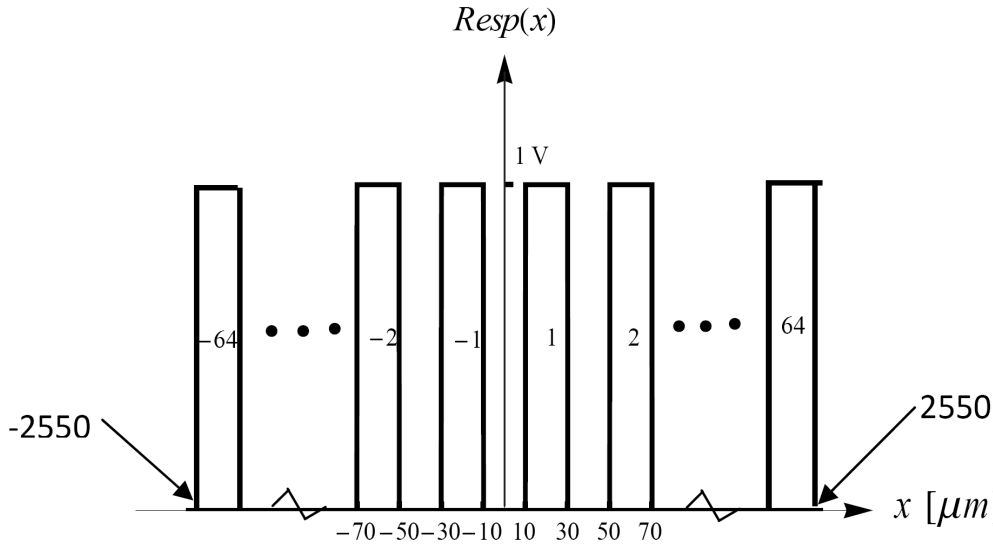


Figure 2.5 Detector response function.

such that the *comb* function is shifted by 20 *μm* and scaled to 40-*μm* center-to-center spacing. The appropriate voltage response model of the array is then

$$Resp(x) = \left[\frac{1}{40} comb\left(\frac{x-20}{40}\right) * rect\left(\frac{x}{20}\right)\right] rect\left(\frac{x}{5120}\right) \tag{2.68}$$

where x is given in micrometers. The *rect* multiplier is the limiting window function that allows only 128 detectors to exist. The response beyond this array is zero due to the definition of the $rect(x)$ function. This type of model can easily be extended to two dimensions because most detector arrays can be described by separable functions.

Example 2.2
The optical transfer function (in the frequency domain) of any infrared or EO sensor with a particular circular limiting aperture can be described by the autocorrelation of a *cyl* function model of the aperture. The autocorrelation of a *cyl* function with unity diameter yields

$$\gamma_{cyl}(\rho) = cyl(\rho) \bigstar\bigstar cyl(\rho) = \frac{2}{\pi}\left[cos^{-1}(\rho) - \rho(1-\rho^2)^{\frac{1}{2}}\right] cyl\left(\frac{\rho}{2}\right) \tag{2.69}$$

where a good approximation has been shown to be

$$\gamma_{cyl}(\rho) \approx 0.25\left[5\, tri(\rho) - 1 + \rho^4\right] cyl\left(\frac{\rho}{2}\right) \tag{2.70}$$

Shifting and scaling of these two functions is performed in the same manner as previously shown. Note that the complex autocorrelation of this function is the same as the autocorrelation because there is no imaginary part or nonzero phase associated with the function. Also, the cross-correlation of two cylinder functions of different diameters is much more complicated than the function described above, and the corresponding analysis and functions can be found in Gaskill [10].

Example 2.3
The electron beam impact on the phosphorus coating of a CRT display causes photoluminescence that is modeled by a Gaussian intensity. The smaller the Gaussian spot diameter, the higher the spatial resolution the CRT provides (because higher spatial frequencies can be passed). The transfer function that describes the passage of spatial frequencies is represented by the Fourier transform of the Gaussian spot. For example, the transfer function of a 0.25-mm spot CRT can be described by

$$\mathcal{FF}\left\{Gaus\left(\frac{r}{0.25^2}\right)\right\} = \frac{1}{0.25^2} Gaus(0.25\rho) \tag{2.71}$$

where r is in millimeters and ρ is in cycles per millimeter.

The three examples shown here provide insight into functional modeling, correlation, and the Fourier transform. None of these examples appears to be difficult,

but a good deal of the analytical work has already been performed and we are using the results. There are not many models within the realm of infrared and EO sensor analysis that are difficult to use or derive. It is still important to understand the basic mathematical concepts; then when modifications are necessary, they can be performed correctly. The mathematics described in this chapter will be used extensively throughout the discussions of sensor analysis techniques in this book.

2.10 Guide to the References

Good overall books on mathematics useful for understanding electro-optics are Andrews [2] and Gaskill [10]. McMahon [15] has an easy-to-read book on complex variables. More detailed treatments of complex variables are given by Saff [21] and Brown [6]. Special functions useful for electro-optics are given by Gaskill [10] and Bracewell [5,4]. Note that the notation for the special functions is not standardized. For example, the meaning of $\delta\delta(x)$ and $\delta_\delta(x)$ differ in Bracewell [5] and Gaskill [10]. Be aware that *sinc* function is defined without π by some authors. As a result the definition for *sinc* used by *Mathematica* [2] differs from the definition used in this text. For the most part, this book adopts the notation of Gaskill [10]. Chapters on convolution can be found in Bracewell [4, 5], Champeney [7], Gaskill [8], and Hsu [11]. With regard to the Fourier integral, some books use $2\,\pi\,f$ in the exponent whereas others use ω and this leads to differences in the Fourier transform normalization factor. Also some books use $+\,j$ in the exponent and others use $-j$ in the exponent when defining the Fourier transform. With regard to Fourier integral notation and the Fourier transform normalization factor, this book agrees with the notation of Gaskill [10] and Bracewell [4,5]. Other books that provide excellent treatments on the Fourier integral are those by Champeney [7], Hsu [14], and Papoulis [16]. Books on probability arranged from easy to read to more detailed and more complex are those by Ash [3], Hsu [13], and Papoulis [16]. Books on noise arranged from easy to read to more detailed and more complex are those by Hsu [13], Cooper [9], Peebles [19], and Davenport [9]. Szirtes [22] has an interesting book that demonstrates how one can use dimensional analysis to derive models. With regard to software, *Mathcad* [S2] is easiest to learn, *MATLAB* [S3] is most commonly used by engineers, and *Mathematica* [S1] exceeds at symbolic manipulation and has strong numerical, graphical, and scientific word processing capability.

2.11 Exercises

2.1 Give $f(x)$ in polar form:
a) $f(x) = 20\ e^{j110°} - 8\ e^{j40°}$
b) $f(x) = 6.1\ x - j\ 3.82$

2.2 Give $f(x)$ in rectangular form:
a) $f(x) = \dfrac{j}{6.3 - j9.7}$
b) $f(x) = j\ x^2 + j\ x + x\ e^{-j45°}$

2.3 Draw the following functions:

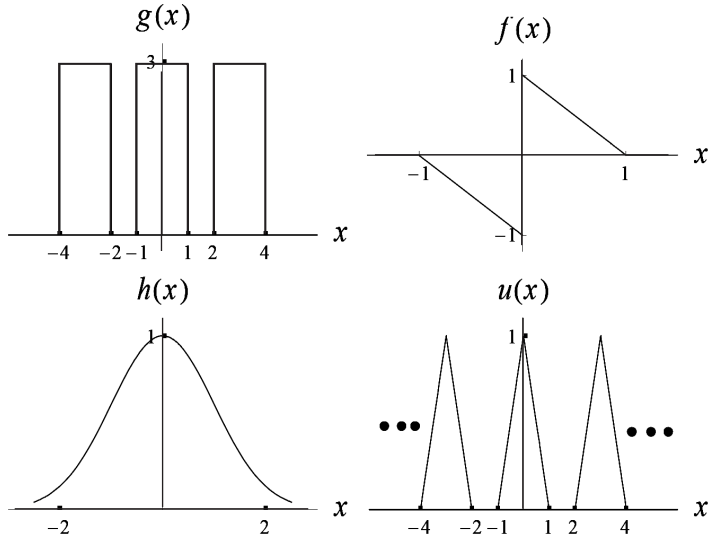
a) $f(x) = 2\,rect\left(\frac{x-3}{5}\right)$

b) $f(x) = Gaus(3x-2)$

c) $f(x) = 4\,rect\left(\frac{x-4}{4}\right) - 2\,rect\left(\frac{x-4}{2}\right)$

d) $f(x) = 4\,tri\left(\frac{x-5}{2}\right) - 2\,tri(x-5)$

2.4 Write a function for the following signals:



2.5 From Exercise 2.4, sketch $v(x) = 4f\left(\frac{2-x}{3}\right)$

2.6 Sketch the following functions:

a) $f(x,y) = rect\left(\frac{x}{2}, \frac{y}{4}\right)$

b) $f(x,y) = sinc\left(\frac{x}{2}, y\right)$

c) $f(x,y) = cyl\left(\frac{\sqrt{x^2+y^2}}{2}\right) - cyl\left(\sqrt{x^2+y^2}\right)$

d) $f(x,y) = somb\left(\sqrt{x^2 + \left(\frac{y}{2}\right)^2}\right)$

e) $f(x, y) = rect(x-2)\delta(y)$

f) $f(x, y) = \delta(x)\ \delta(2-y)$

2.7 Sketch the following functions:

a) $f(x) = rect(x) * rect(x-2)$

b) $g(x) = \delta\delta(x) * \delta\delta(x)$

c) $h(x) = step(x) * rect\left(\frac{x}{4}\right)$

d) $u(x) = rect\left(\frac{x}{4}\right) * rect\left(\frac{x}{3}\right)$

e) $v(x) = \frac{1}{4} comb\left(\frac{x}{4}\right) * rect\left(\frac{x}{2}\right)$

2.8 Perform the following convolutions or correlations and sketch the results:

a) $f(x,y) = rect(x,y) ** rect\left(\frac{x}{2}, \frac{y}{2}\right)$

b) $rect(x, y) ** \delta(x - 3, y - 2)$

c) $f(x,y) = sinc\left(\frac{x}{2}, \frac{y}{4}\right) ** \delta\delta\left(\frac{x}{4}\right)$

d) $f(x, y) = cyl(2r) \bigstar\bigstar cyl(2r)$

2.9 Find the Fourier transform of the following functions:

a) $f(x) = |x|$

b) $f(x) = e^{-x^2}$

c) $f(x) = rect\left(\frac{x}{3}\right)$

d) $f(x) = rect(x - 3)\ e^{j2\pi x}$

e) $f(x) = rect(x - 1) * rect\ (x)$

f) $f(x) = sinc(4x) * sinc\ (2x)$

g) $f(x) = \frac{1}{2} comb\left(x - \frac{1}{2}\right) * tri(x)$

2.10 Find the Fourier transform of the following 2D functions:

a) $f(x,y) = rect\left(\frac{x}{2}, y\right)$

b) $f(x,y) = \frac{1}{5} comb\left(\frac{x}{5}\right) ** rect\left(\frac{x}{2}, y\right)$

c) $g(r) = cyl\left(\frac{r}{2}\right) \bigstar\bigstar cyl\left(\frac{r}{2}\right)$

2.11 Show that:

a) $lim_{b\to 0} \frac{1}{|b|} Gaus\left(\frac{x}{b}\right)$ is a reasonable representation of $\delta(x)$.

b) $\mathcal{F}\{e^{j2\pi ft}\} = \delta(f)$ and use this result to show that (2.28) and (2.29) are reasonable.

2.12 Show that (2.69) is correct.

2.13 Show that $\mathcal{FF}\left\{Gaus\left(\frac{r}{a^2}\right)\right\} = \frac{1}{a^2} Gaus(a\,r)$.

2.14 Assume A={a,b,c,d,e,f}, B={e,f,g,h,i,j}, C={g,h,i}, D={j,k} and that the outcomes a, b, …, k are equally likely. Calculate:

a) $P(A)$, $P(B)$, $P(C)$, $P(D)$

b) $P(A \cup B)$, $P(\mathrm{A} \cup \mathrm{C})$, $P(\mathrm{A} \cup \mathrm{B} \cup C)$
c) $P(A \cap B)$, $P(\mathrm{A} \cap \mathrm{C})$
d) $P(\mathrm{C}|A)$, $P(C|B)$

2.15 Show:
a) If $g(x) = f(x)*h(x)$, then $g(x - x_0) = f(x - x_0) * h(x)$.
b) $f(x) *\delta(x) = f(x)$

References

[1] Andrews, L., *Special Functions for Engineers and Applied Mathematicians*, New York City, NY: McMillan, 1985.

[2] Andrews, L., and R. Phillips, *Mathematical Techniques for Engineers and Scientists*, Bellingham, WA: SPIE Press, 2003.

[3] Ash, C., *The Probability Tutoring Book*, Piscataway, NJ: IEEE Press, 1993.

[4] Bracewell, R. N., *Two-Dimensional Imaging*, NJ: Prentice Hall, New Jersey, 1995.

[5] Bracewell, R. N., *The Fourier Transform and Its Applications*, *Third Edition*, New York City, NY: McGraw-Hill, 2000.

[6] Brown, J., and R. Churchill, *Complex Variables and Applications*, New York City, NY: McGraw-Hill, 2008.

[7] Champeney, D. C., *Fourier Transforms and Their Physical Applications*, New York City, NY: Academic Press, 1973.

[8] Cooper, G., and C. McGillem, *Probabilistic Methods of Signal and System Analysis, 4th Edition*, New York City, NY: McGraw-Hill, 1999.

[9] Davenport, W., and W. Root, *An Introduction to the Theory of Random Signals and Noise*, New York City, NY: McGraw-Hill. 1987.

[10] Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics*, New York City, NY: John Wiley & Sons, 1978.

[11] Goodman, J. W., *Introduction to Fourier Optics*, San Fransisco, CA: McGraw-Hill, 1968.

[12] Hayt, W., and J. Kemmery, *Engineering Circuit Analysis*, New York City, NY: McGraw-Hill, 1978.

[13] Hsu, H., *Probability, Random Variables and Random Processes*, New York City, NY: McGraw-Hill, 1997.

[14] Hsu, H., *Applied Fourier Analysis*, New York City, NY: Harcourt Brace Jovanovich, 1984.

[15] McMahon, D., *Complex Variables DeMystified*, New York City, NY: McGraw-Hill, 2008.

[16] Papoulis, A., *The Fourier Integral and Its Applications*, New York City, NY: McGraw-Hill, 1962.

[17] Ross, S. L., *Introduction to Communication Systems*, New York City, NY: Wiley, 1966.

[18] Papoulis, A., *Systems and Transforms with Applications in Optics*, Malabar, FL: Robert Kreiger Publishing Co., 1981.

[19] Papoulis, A., and S. Pillai, *Probability, Random Variables and Stochastic Processes*, New York City, NY: McGraw-Hill, 2002.

[20] Peebles, P., *Probability, Random Variables and Random Signal Principles*, New York City, NY: McGraw-Hill, 2000.

[21] Ross, S. L., *Introduction to Communication Systems*, Hoboken, NJ: Wiley, 1966.

[22] Saff, E. B., and A. D. Snider, *Fundamentals of Complex Analysis for Mathematics, Science, and Engineering*, Upper Saddle River, NJ: Prentice Hall, 1976.

[23] Stremler, F. G., *Introduction to Communication Systems*, Reading, MA: Addison-Wesley, 1982.

[24] Szirtes, T., *Applied Dimensional Analysis and Modeling*, New York City, NY: McGraw-Hill, 1998.

Software

[S1] *Mathematica*, http://www.wolfram.com/products/mathematica/index.html.
[S2] *Mathcad*, http://www.ptc.com/products/mathcad.
[S3] *MATLAB*, http://www.mathworks.com.

CHAPTER 3

Linear Shift-Invariant Systems

Linear operators and processes are a key part of an engineer's or scientist's formal education in analytical mathematics. Integration, derivatives, and Fourier transforms are all linear operations. As a starting point, most systems are assumed to be linear, and a model is developed using the mathematics of Chapter 2. In reality, most systems are approximately linear over a restricted region of their application. In this chapter we describe some general properties of linear shift-invariant (LSI) systems and demonstrate how they may be applied to imaging systems.

As a preview of the material covered in this chapter, consider the two systems depicted in Figure 3.1. A simple *RC* circuit is shown with input and output functions of the independent variable of time. Also shown is an imaging system where the independent variables are 2D space. Both the circuit and the imaging system are also shown in the frequency domain where the circuit frequency is in cycles per second, or Hertz, and the imaging system frequency is given in cycles per meter (or cycles per milliradian). These systems are both modeled as LSI and are consequently characterized by the *impulse response function h*., which is the system response to an input with a short duration compared to the time response of the system. For the circuit, $h(t)$ is the response to a voltage spike of infinitesimal temporal duration; for the imaging system we are interested in, the response $h(x,y)$ is to a point of light on the input. The output of the circuit gives a voltage impulse response, and the output of the imaging system gives the spatial impulse response (equal to the blur image of the point source). It will be shown that once the impulse response is determined, the output of the system with any input function is simply the input function convolved with the impulse response.

Equivalently we can look at the response of an LSI system as a function of input *frequency* (rather than time or space). For example, in an imaging system we might want to know the degree to which different wavelengths of light are attenuated by the imager. In the frequency domain, the spectrum of the spike is a uniform value (1D transform for circuits and 2D transform for imaging systems). This uniform input spectrum is multiplied by the circuit or imaging system transfer function to give an output spectrum. This output spectrum is the transfer function of the system. With circuits, this is simply the transfer function with a magnitude and phase. With imaging systems, the transfer function is described by a modulation transfer function that is similar to the magnitude of the circuit transfer function. Once the transfer function is determined, the output spectrum, given any input spectrum, can be found by multiplying the input spectrum by the transfer function.

In summary, there are two methods for LSI analysis: (1) the time or space domain where the input function is convolved with the impulse response to give an

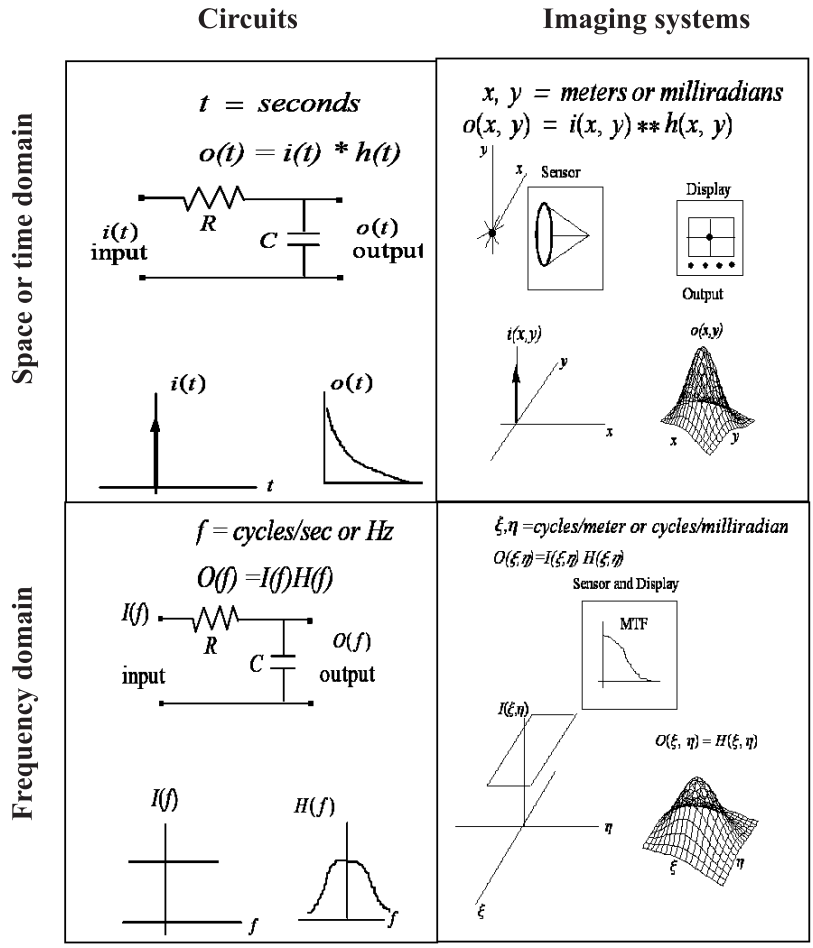


Figure 3.1 Circuit and imaging system LSI analysis.

output, and (2) the frequency domain where the input spectrum is multiplied by the system transfer function to give an output spectrum. The transform of the impulse response gives the transfer function, and the inverse transform of the transfer function gives the impulse response. The LSI approach to modeling both spatial and frequency domain system responses is covered in this chapter.

Finally, we should mention that in the past 5 to 10 years, many infrared systems have migrated from scanning systems to staring systems, where the staring systems are undersampled. The infrared detectors are large (on the order of 20 μm) compared to the optical blur spot. That is, the sample spacing is not small enough to reconstruct the continuous image signal formed by the optics. Shannon's criterion is not satisfied and the resulting image is undersampled. These systems are still very good imaging systems, however, they are not LSI systems. They are certainly linear, but they are not shift invariant. While the majority of the material in this book assumes LSI systems, it is useful to understand when a system is not LSI and some possible ways to treat these systems in design and analysis. Sampling is discussed in detail in Sections 3.7, 3.8, and 3.9.

3.1 Linear Systems

Linearity requires two properties [5, 6], superposition and scaling. Many scientists and engineers study temporal systems such as circuits, where the independent variable is time t. There is no difference in the mathematics if the spatial variables, x and y, are the independent variables instead. To illustrate this concept, we use a 1D analysis. Consider the input $i(x)$ to a system modeled with the operator $L\{\}$, producing the output

$$o(x) = L\{i(x)\} \tag{3.1}$$

For separate inputs $i_1(x)$ and $i_2(x)$ the system is said to provide a superposition characteristic if

$$o(x) = L\{i_1(x) + i_2(x)\} = L\{i_1(x)\} + L\{i_2(x)\} \tag{3.2}$$

Such a system behaves as though the inputs were present by themselves and the output is simply the sum of the individual responses. The property of superposition is important for analysis of linear systems. As will be shown, this property allows us to determine the output for a simple input and then use this result to predict the output for much more complicated input signals, described as a sum of input functions.

For the systems we often deal with, the output $o(x)$ is analytically obtainable by using a complex exponential (i.e., sinusoidal) input. The system output for a set of sinusoidal input functions is sometimes termed the sine wave response. For optical systems, the relationship between the sine wave response and sinusoidal input is sometimes called the *modulation transfer function* (MTF) [8]. The frequency for sinusoids of spatial variables x and y, for example, is expressed in cycles per millimeter or some other appropriate scaled dimension. The response to a delta function is called the *impulse response*. In an optical system, the response can also be called a *point spread function* (psf).

We can observe the psf easily by looking at the stars on a clear night. Virtually all of the visible stars will appear to be about the same size, in spite of the vast differences in their actual sizes and distances from earth. What is really being sensed is the psf of the eye because the stars are too small to be resolved. That is, the stars are point sources, or delta functions, for the human eye. In mechanical systems, a sudden input (or impulse) might be achieved by striking the system. Light flashes viewed by video cameras (as frequently seen at nighttime college football games) provide impulses both in time and space.

The second property required for a system to be linear is the scaling property. Briefly, if the system is linear, doubling the input will only result in a doubling of the output. "Doubling" is simply an arbitrary chosen scale factor to illustrate the property. Whatever scale factor is applied to the input, the same scale factor must result at the output. Using our previous notation, this is expressed in equation form as

$$o(x) = L\{a\, i(x)\} = a\, L\{i(x)\} \tag{3.3}$$

A common feature in infrared and EO systems that defeats this property is the presence of automatic gain control (AGC). If the gain of the system adjusts

depending on the input, then the scaling property does not hold for the system. EO test engineers must be aware of any AGC function in the sensor electronics. For tests requiring linearity, the AGC must be controlled, either by disabling it or by presenting a very bright input in a part of the field of view (FOV). The bright input drives the AGC to its minimum gain condition and holds it there so that responses to test objects obey the scaling property. Also, saturation must be avoided or the scaling property will not hold. Saturation occurs when the input signal is large enough to cause the output signal to reach its maximum. Under this condition, changes seen on the input signal are not seen on the output signal.

3.2 Shift Invariance

Shift invariance [3, 7] in a temporal system implies that the system is not changing with time. The response to an input applied now is the same as the response to that input applied an hour from now. It will simply be shifted in time by an hour. In temporal systems, this property is often referred to as time invariance. Shift invariance in an optical system implies that an input at one point in the FOV produces a response identical to that of the same input at a different location in the FOV, just shifted to a different location. In optical systems, this is frequently a marginally fulfilled condition. Most optical systems perform better "on axis," near the axis of symmetry of the optical elements, than they do "off axis" (although sometimes the optical elements themselves are off-axis elements). Thus, shift invariance is considered a first-order assumption for most optical systems. It is approximately true over small regions in the FOV.

The property of shift invariance can be stated in equation form. Let $o_1(x)$ represent the output resulting from an input located at x; that is, $i(x)$. Then, the output $o(x)$ in response to the same input shifted to x_0 is

$$o(x) = L\{i(x - x_0)\} = o_1(x - x_0) \tag{3.4}$$

for shift invariance to hold.

Shift invariance allows the techniques of linear systems analysis to be applied to infrared and EO systems. In effect, the system responds to identical inputs in the same way regardless of where they occur in the FOV. Without shift invariance, the system would require characterization at every location within its FOV. Characterization at a single location is sufficient for a LSI system.

3.3 Basics of LSI Systems

Sections 3.1 and 3.2 revealed the basic characteristics of LSI systems. An arbitrary input is considered to be composed of a sum (i.e., integral) of appropriately weighted and shifted input functions (sine wave, delta functions, and so on). This means, for example, that the response (output) of the system to a single delta function can then be used, considering superposition, to determine the output of the system to this arbitrary input. Thus, all we need to completely characterize the input/output behavior of an LSI system is its impulse response.

We begin with a brief review of delta functions. The delta function, $\delta(x)$, has already been discussed in Chapter 2 and is characterized by its mathematical properties, namely, infinite height, zero width, and an area of one:

$$\int_{\infty}^{\infty} \delta(x)\, dx = 1$$

It is perhaps clearer to picture the delta function as the limit of a process of shrinking the width and increasing the height so as to retain unit area. Convenient functions to shrink in width are Gaussian, rectangle, and *sinc* functions, although others will also yield delta functions in the limit of zero width. The infinite height is sometimes troublesome to the practical-minded analyst or engineer. In the real world, only the integration properties of the delta functions are used. The most fundamental of these is the sifting property (2.6b). It states that multiplying a function by a delta function and integrating across the delta function sifts out the value of the function at the location of the delta function. Recall from Chapter 2 that the sifting property was written as

$$\int_{x_1}^{x_2} g(\alpha)\delta(\alpha - x_0)\, d\alpha = g(x_0) \quad x_1 < x_0 < x_2 \tag{3.5}$$

Briefly stated, we multiplied by a delta function located at x_0, integrated across the delta function, and obtained the value of $g(x)$ at the location of x_0 [i.e., $g(x_0)$].

We can use this property to represent an *arbitrary* function $f(x)$ as the integral (superposition) of appropriately weighted delta functions:

$$f(x) = \int_{-\infty}^{\infty} f(\alpha)\delta(\alpha - x)\, d\alpha \tag{3.6}$$

Delta functions therefore provide a convenient way of describing an arbitrary system input as a superposition of a very simple system input. By (3.2), (3.3), and (3.4), if we know the system response for a single delta function, we can derive the response to an arbitrary function given by (3.6). To see that this is true, let the input/output for the LSI system $L\{\}$ be given by

$$o(x) = L\{i(x)\} \tag{3.7}$$

The input can be written as a sum of weighted and appropriately located delta functions:

$$i(x) = \int_{-\infty}^{\infty} i(\alpha)\delta(x - \alpha)\, d\alpha \tag{3.8}$$

Then the output is

$$o(x) = L\left\{\int_{-\infty}^{\infty} i(\alpha)\delta(x - \alpha)\, d\alpha\right\} \tag{3.9}$$

The linear operator operates only on functions of x, so the operator can be moved inside the integral to yield

$$o(x) = \int_{-\infty}^{\infty} i(\alpha) L\{\delta(x-\alpha)\}\, d\alpha \tag{3.10}$$

where $L\{\delta(x-\alpha)\}$ is the response of the system to a delta function located at $x = \alpha$. The shift-invariance property indicates that the response to an impulse at α is the same as the response to an impulse at o, just shifted to α. Letting $h(x)$ represent the system response to an impulse at $x = o$, the response to an impulse at α is $h(x-\alpha)$. Substitution yields the desired result:

$$o(x) = \int_{-\infty}^{\infty} i(\alpha) h(x-\alpha)\, d\alpha \tag{3.11}$$

This is just the convolution of the input with the system's impulse response, and (3.11) is referred to as the *convolution integral*. Using the notation introduced in the previous chapter we may also write (3.11) as follows:

$$o(x) = i(x) * h(x) \tag{3.12}$$

3.4 Impulse Response

The concept of the impulse response can best be shown through an example. Consider the resistor-capacitor (RC) circuit shown in Figure 3.2(a). The impulse response of the circuit can be found using circuit theory [12] as

$$h(t) = \frac{1}{RC} e^{-t/(RC)} \text{ for } t \geq 0 \tag{3.13}$$

where the positive time constant of the circuit is given by the product RC [sec]. The graphic representation is shown in Figure 3.2(b). We are interested in determining the output voltage $v_0(t)$, given the *step* input function shown in Figure 3.2(c). The output corresponds to the charging of the capacitor through the resistor given some applied dc voltage $v_i(t)$. The convolution of the input with the impulse response gives the output. Recall from Chapter 2 that the convolution involves the flip, slip, multiply, and integrate process, where $h(t)$ is the function that is flipped (remember that convolution is commutative). The flipped function and the stationary function plotted are shown in Figure 3.2(d). These two functions are multiplied and integrated. The integration from $-\infty$ to ∞ can be broken into three sections: $(-\infty, 0)$, $(0, t)$, and (t, ∞). Note that multiplication of the two functions (point by point) yields a zero product unless $t \geq 0$ as shown in Figure 3.2(e). The only section that yields a nonzero product of the two functions is $(0, t)$. Therefore, the convolution integral becomes

$$v_0(t) = \int_0^t (1)\left(\frac{1}{RC}\right) e^{-(t-\alpha)(RC)}\, d\alpha \tag{3.14}$$

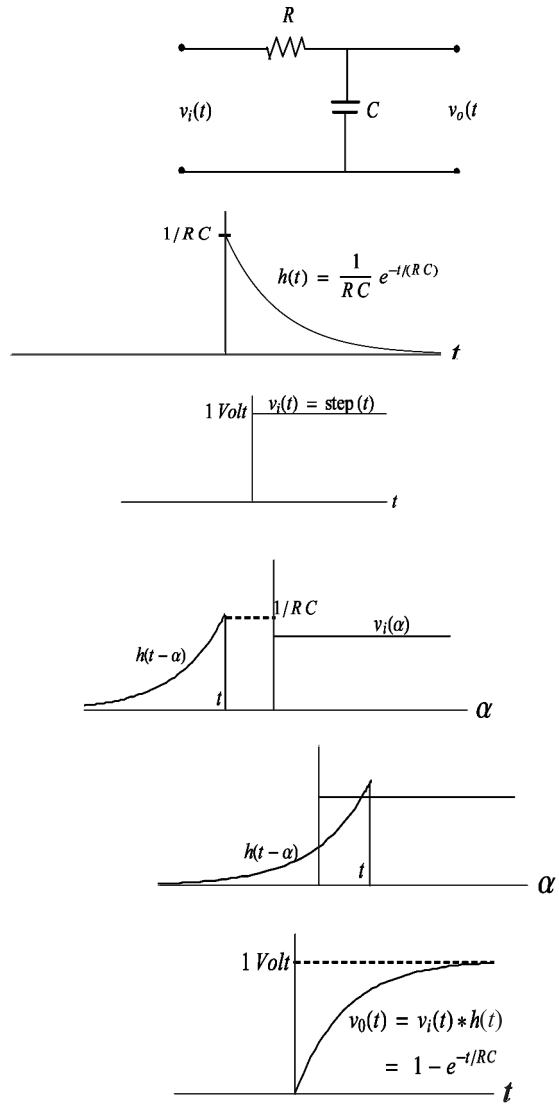


Figure 3.2 Impulse response example.

Evaluation of the integral gives

$$\begin{aligned} v_0(t) &= \frac{1}{RC} e^{-\frac{t}{RC}} \int_0^t e^{\frac{\alpha}{RC}}\, d\alpha \\ &= e^{-\frac{t}{RC}} \left[e^{\frac{t}{RC}} - e^0 \right] = 1 - e^{-\frac{t}{RC}} \end{aligned} \tag{3.15}$$

for arguments greater than or equal to 0. The output is shown in Figure 3.2(f). The output appears to be reasonable in that (1) the voltage of the capacitor approaches the source voltage (i.e., 1V) and (2) the rate at which the capacitor

voltage approaches the source voltage is inversely proportional to the product of the resistance and capacitance.

It is worth showing a simple impulse response example of an optical system, where the output can be determined for some well-defined input. Consider the imaging system shown in Figure 3.3. The optical system shown by the box could be a series of lenses or a single lens that reimages two points of light onto an image plane. The input object plane is denoted by x_{ip}, y_{ip} and the image plane is denoted x_{oi}, y_{oi}. The two points of light are small enough in spatial extent to be unresolvable by the system and, therefore, can be considered impulses. For this example, we assume that the light is monochromatic (single frequency) and that the system can be considered spatially coherent. Coherence is discussed in Chapter 4. The important point here is that we want to find the electromagnetic field output for the given input. From Chapter 2, the shifting and scaling of the even impulse pair is given by

$$\frac{1}{|a_0|}\delta\delta\left(\frac{x_0}{a_0}\right) = \delta(x_0 + a_0) + \delta(x_0 - a_0) \tag{3.16}$$

Therefore, the electromagnetic field for the two points of light located at $x_0 + a_0$ from the optical axis in the input plane can be represented with an even impulse pair function:

$$U_0(x_0, y_0) = \frac{1}{|a_0|}\delta\delta\left(\frac{x_0}{a_0}\right) \tag{3.17a}$$

The impulse response or psf for a diffraction limited optical system in this example has been determined to be [see (4.38)]:

$$h(x_i, y_i) = psf(x_i, y_i) = \frac{\pi}{4b^2} somb\left(\frac{r}{b}\right) \tag{3.18}$$

where b is $\lambda f/d$. Here λ is the wavelength of the monochromatic light, d and f are the diameter and focal length of the optical lens, and $r = \sqrt{x_i^2 + y_i^2}$. Note from (3.18) that r has units of length. If the impulse response was a delta function, then the field in the image plane would be

$$U_{ideal} = (x_i, y_i) = \frac{1}{|a_i|}\delta\delta\left(\frac{x_i}{a_i}\right) \tag{3.17b}$$

where $\pm a_i$ are the positions on the x_i axis of the point source according to geometric optics. Recalling the shift invariance property of (2.13) and the delta function convolution property of (2.14), the field in the image plane is

$$\begin{aligned} U_i(x, y) &= U_{ideal}(x_i, y_i) * h(x_i, y_i) \\ &= \frac{\pi}{4b^2}\left[somb\left(\frac{\sqrt{(x_i + a_i)^2 + y_i^2}}{b}\right) + somb\left(\frac{\sqrt{(x_i - a_i)^2 + y_i^2}}{b}\right)\right] \end{aligned} \tag{3.19}$$

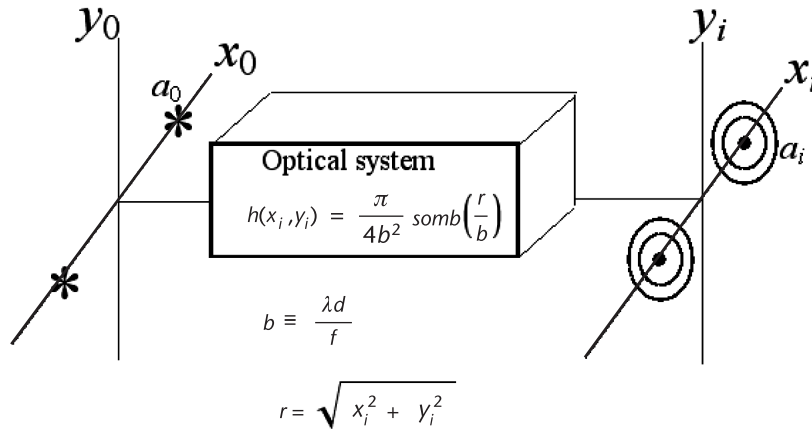


Figure 3.3 Optical system response.

It is important to convince yourself that this is correct. While the result is two shifted and scaled impulse responses, the tools of linear systems are shown to work in the optical case. Making the substitution $b = \lambda f/d$ shows that the diameter of the *somb* function (sometimes called the Airy disc after the gentleman who originally derived it) increases with an increased wavelength or a decreased optical lens diameter (or limiting aperture). This simple example contains several important concepts that occur again and again.

The field magnitude derived here is the system output. It is not sensed by vision or by optical detectors. The square of the field magnitude is defined as the irradiance and is the parameter that is actually sensed in imaging systems. It corresponds to optical power. Also, most infrared and EO systems are not monochromatic and they operate in a "nonspatially coherent" environment. The usual noncoherent impulse response is proportional to the square of the coherent impulse response (i.e., the *somb* function). This noncoherent impulse response is the psf. The relationship causes noncoherent systems to be LSI systems of the irradiance variable instead of the electromagnetic field quantity. *In simple terms, the input irradiance is convolved with irradiance the psf to determine system output irradiance.*

The impulse response, or psf, of an entire system is the convolution of the individual impulse responses of the optics, detectors, electronics, processor, display, and, if appropriate, human vision. This is due simply to the fact that the output of each component is the input of the next. The result is a more Gaussian-like system psf as shown in Figure 3.4(a). The x-axis cross section is shown in Figure 3.4(b).

It is sometimes difficult to measure the psf of an infrared and EO system owing to the difficulty in creating a point source experimentally. An alternative is to instead measure the *edge response*, of the system. A long, straight target is uniformly illuminated from behind, and the sensor image transition from dark to bright is measured as shown in Figure 3.4(c). The x-axis cross section is shown in Figure 3.4(d). Mathematically, the edge response is the response of an imaging system to a *step* function input and is a quantity that is easily measured by experiment. It can be shown that the relationship between the impulse response $h(x,y)$ (psf) and the edge response $e_x(x)$ is [5]

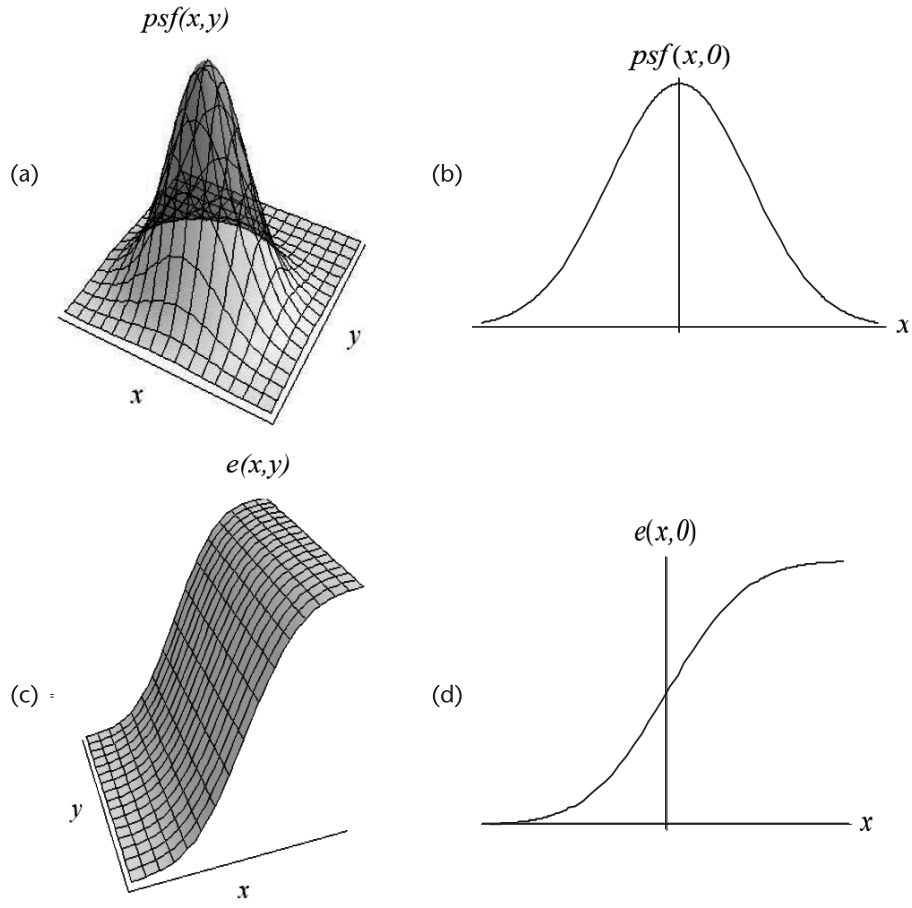


Figure 3.4 The psf and edge response.

$$
\begin{aligned}
e_x(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\alpha,\beta)\, step(x-\alpha)\, d\alpha\, d\beta \\
&= \int_{-\infty}^{\infty} h \int_{-\infty}^{\infty} \left[(\alpha,\beta)\, d\beta\right] step(x-\alpha)\, d\alpha \\
&= \int_{-\infty}^{\infty} l(\alpha)\, step(x-\alpha)\, d\alpha \\
&= \int_{-\infty}^{x} l(\alpha)\, d\alpha
\end{aligned}
\tag{3.20}
$$

The inner integral is referred to as the *line spread function* (lsf) and is obtained by integrating the psf over one of the spatial variables. From (3.20) it is apparent that the lsf is the spatial derivative of the edge response (in the direction normal to the blade edge); that is,

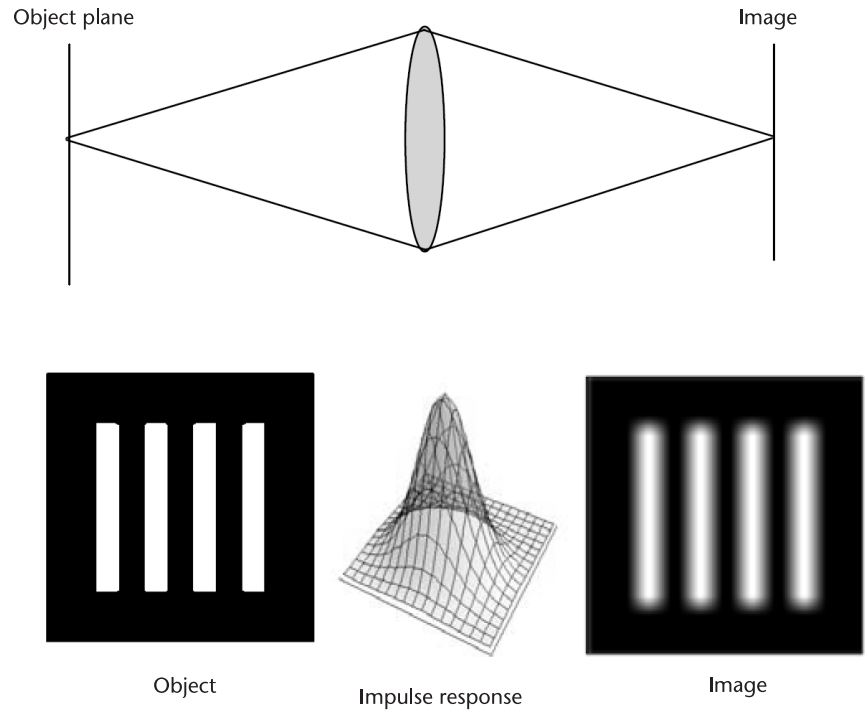


Figure 3.5 Four-bar target, imaging system, and image.

$$l(x) = \frac{d\, e_x(x)}{dx} \tag{3.21}$$

and provides a straightforward method of obtaining the impulse response of the system in practice. One can, for example, approximate $h(\alpha, \beta)$ with a 2D Gaussian distribution, yielding a lsf that is also Gaussian. The edge response will therefore be an "error function" [11], whose parameters can be estimated from the acquired data. The psf is then also completely characterized by the estimated parameters.

We consider a final example to firm up the impulse response concept. It is typical, in imaging systems, to evaluate the performance of an imaging system by placing a four-bar target in front of the imager as the object and to evaluate the quality of the four-bar target image as shown in Figure 3.5. The four-bar target is simply a transparency with four rectangular holes so that light (visible or infrared) can pass. The image of the four-bar target is always a degraded rendition of the original object. The system is the optical system shown where the transparency is the input and the image is the output.

Because the system is a LSI system, from (3.12) we know that the output is the input convolved with the system impulse response (or psf). Figure 3.5 shows the output that is the convolution of the object with the impulse response.

3.5 Transfer Function

The system transfer function [7, 3, 12] defines the system frequency throughput. In temporal systems, a wide transfer function allows high frequencies through the

system. High temporal frequencies correspond to very rapid changes in some parameters such as voltage or current with respect to time. In space, a wide transfer function allows high spatial frequencies through a system. High spatial frequencies correspond to the irradiance of an object changing dramatically over short distances in space. For example, leaves on a tree produce higher spatial frequency signals than the trunk of the tree.

It is usually convenient to evaluate the response of LSI systems to some input in the frequency domain. In infrared and EO systems, the input spectrum is determined by the Fourier transform of the input irradiance. The input spectrum is then multiplied by the system transfer function to determine the output spectrum. The inverse transform of the output spectrum yields the resulting system output image. The analysis in the space domain requires the input space to be convolved with the system impulse response to produce a spatial output (an image). Note that the duality of convolution in the space domain corresponds to multiplication in the frequency domain. Also, the transfer function is simply the Fourier transform of the impulse response.

The transfer function determines the weighting of the output spectrum relative to the input spectrum. Consider the circuit mathematically defined in Figure 3.6(a). The transfer function of the circuit is shown in Figure 3.6(b). To determine the output for given inputs, the input spectrum is multiplied by the transfer function. This means that each frequency component of the input is assigned a weighting factor as determined by the transfer function. The result is then inverse-transformed to yield the output. Consider the inputs given in Figures 3.6(c–e). The Fourier transform of a constant 1V input, a dc signal, is an impulse located at the origin (see Table 2.4). This impulse is multiplied by the unity throughput of the transfer function for dc signals. The output is the inverse transform of the unaltered impulse (i.e., a 1V constant signal). The second signal, shown in Figure 3.6(d), is a cosine function of 50 Hz. The transform of this signal is an even impulse pair where the two impulses are located at ±50 Hz. Note from Figure 3.6(b) that the transfer function has an attenuation of 0.5 here, so the multiplication of the impulse pair by the signal spectrum yields a signal of one-half the amplitude. The output is inverse-transformed to give the temporal signal shown. Finally, a 100-Hz signal is input to the system and is multiplied by the transfer function weighting of zero. Therefore, 0V is seen on the system output as a result of this input. What is the output of this system if the input is a 100-Hz oscillation on top of a 50-Hz cosine wave with a dc offset of 1V? Remember the principle of superposition.

In this chapter, we limit our optical transfer function discussion to broadband noncoherent optical systems (i.e., the system is linear with irradiance, not electromagnetic field strength). A more general discussion is reserved for Chapter 4. Consider the space domain of Figure 3.7. The system can be characterized by providing an impulse, or point-source, irradiance to the input of the optical system. The output of the system is simply an image of the impulse response, or psf, because the convolution of any function with an impulse is that function. In the frequency domain, the transform of the impulse function is a uniform frequency distribution across the spectrum. The system output spectrum is the uniform spectrum (i.e., a constant 1) multiplied by the system transfer function. Therefore, the output is simply the system transfer function. Note that the transfer function is simply the

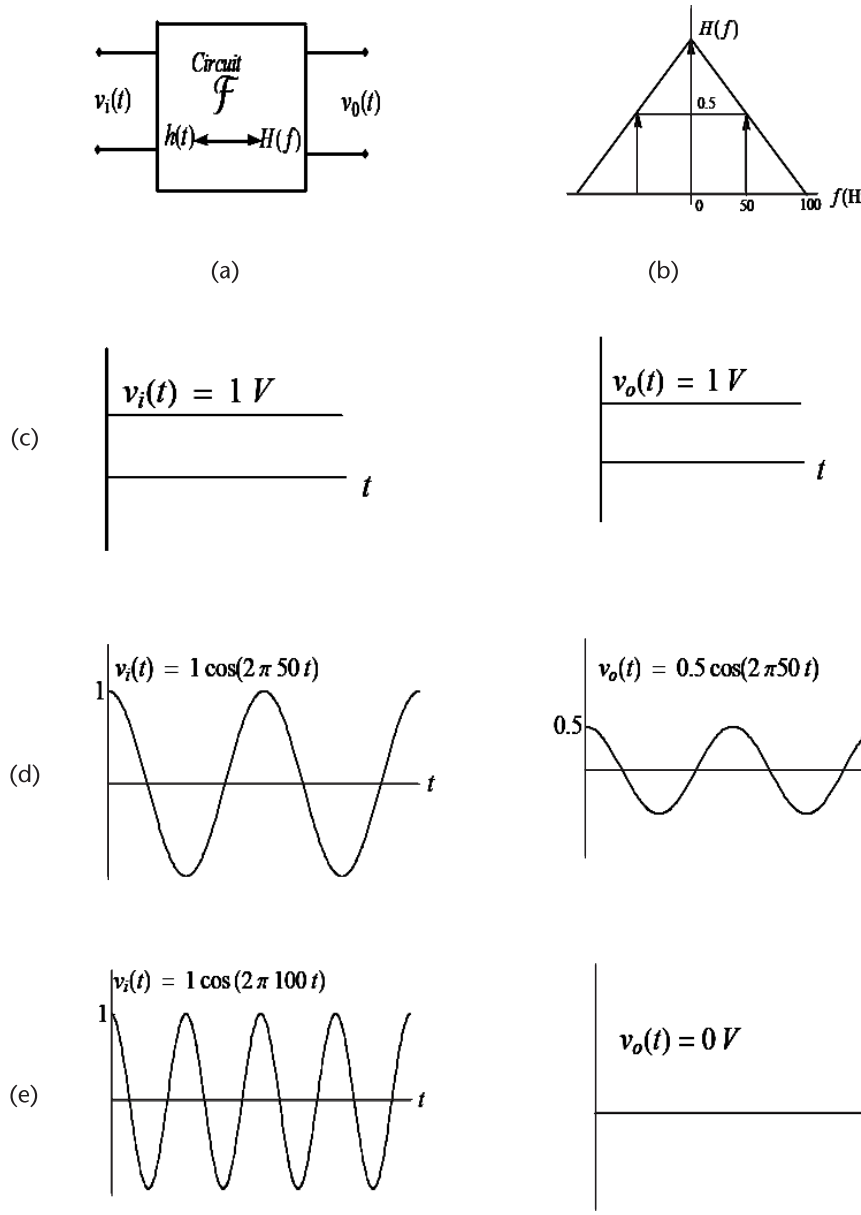


Figure 3.6 Transfer function example.

transform of the impulse response and that narrow impulse responses correspond to wide transfer functions, and vice versa.

The transfer function of a noncoherent system is called the *noncoherent optical transfer function* (OTF). The OTF is complex and has both magnitude and phase. The magnitude is called the *modulation transfer function* (MTF) and the phase is called the *phase transfer function* (PTF). The MTF is often a circularly symmetric function, so it is plotted as a 1D function. The OTF is the Fourier transform of a real function (the psf), so the OTF is Hermitian (the MTF is even and the PTF is odd). The PTF can be ignored if the psf is symmetrical, because the PTF goes to

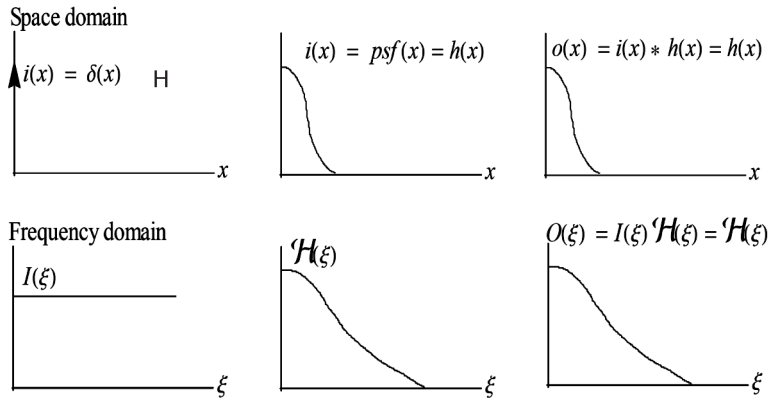


Figure 3.7 LSI optical systems.

zero. Ideally, one could use a point source to obtain a psf and then obtain the OTF by transforming the psf. Practically, this approach can suffer from signal-to-noise limitations so other approaches, such as the edge response, are used.

3.6 System PSF and MTF Versus Component PSF and MTF

The system impulse response is the response of a system to an impulse:

$$h(x,y) = L\{(\delta(x,y)\} \tag{3.22}$$

The output of the system is the convolution of the input scene with the impulse response of the system. That is,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(\alpha,\beta)h(x-\alpha,y-\beta)\,d\alpha\,d\beta = i(x,y)**h(x,y) \tag{3.23}$$

where ** denotes the 2D convolution. The impulse response of the system $h_{system}(x,y)$ is commonly called the psf of the imaging system. The system impulse response is comprised of component impulse responses. Each of the components in the system contributes to the blurring of the scene. In fact, each of the components has an impulse response that can be applied in the same manner as the system impulse response. The blur attributed to a component may be comprised of a few different physical effects. For example, the optical blur is a combination of the diffraction and aberration effects of the optical system. The detector blur is a combination of the detector shape and the finite time of detector integration as it traverses the scene. In a linear shift invariant system the point spread function of the system is the convolution of the individual impulse responses:

$$h(x,y) = h_{atm}(x,y)**h_{opt}(x,y)**h_{det}(x,y)**h_{elec}(x,y)**h_{disp}(x,y)**h_{eye}(x,y) \tag{3.24}$$

so that the total blur, or system psf, is a combination of the component impulse responses.

Remember that the system described above is valid for LSI systems. This analysis technique is a reasonable description for continuous and well-sampled imaging systems. It is not a good description for an under sampled imaging system. These sampled imaging systems do satisfy the requirements of a linear system, but do not follow the shift-invariance property. We describe the sampling nature of these systems later in the chapter. The representation of sampled imaging systems is a modification to the above approach. In this chapter, we continue to develop the foundation of linear systems and Fourier analysis methods applied to imaging systems.

For completeness, we take the *spatial domain* linear systems model and convert it to the *spatial frequency domain*. Given that x and y are spatial coordinates in units of milliradians, the spatial frequency domain has independent variables of ξ and η in cycles per milliradian. A spatial input or output function is related to its spectrum by the Fourier transform

$$F(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(\xi x+\eta y)}\,dx\,dy \tag{3.25}$$

The inverse Fourier transform converts an image spectrum to a spatial function

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\xi,\eta)e^{j2\pi(x\xi+y\eta)}\,d\xi\,d\eta \tag{3.26}$$

The properties and characteristics of the Fourier transform are reviewed in Chapter 2. Here we make use of the fact that the Fourier transform of a convolution integral yields a product. Therefore, the spatial convolution described in (3.23) results in a spectrum of

$$O(\xi,\eta) = I(\xi,\eta)H(\xi,\eta) \tag{3.27}$$

that is, the output spectrum is related to the input spectrum by the product of the Fourier transform of the system impulse response. The Fourier transform of the system impulse response is called the *transfer function* of the system. In fact, each of the component impulse responses given in (3.24) has a component transfer function:

$$O(\xi,\eta) = I(\xi,\eta)H_{atm}(\xi,\eta)H_{opt}(\xi,\eta)H_{det}(\xi,\eta)H_{elec}(\xi,\eta)H_{disp}(\xi,\eta)H_{eye}(\xi,\eta) \tag{3.28}$$

Note that the system transfer function is the product of the component transfer functions

3.7 Spatial Sampling

It is common to represent a function $f(x, y)$ by an array of sampled values. Not only is this the case with infrared and EO systems, but it is also true for synthetic aperture radar (SAR), magnetic resonance imaging (MRI), and other imaging sensors. It is well known that the spatial samples must be taken sufficiently close to each other to obtain an accurate representation of the function. Also, for band-limited functions, the original function can be reconstructed exactly with the sampled array

information provided that the spacing is within particular limits. Band-limited functions are those whose corresponding Fourier transforms are nonzero over some finite frequency band. All frequency components outside the finite frequency band are zero. The reconstruction theory of these band-limited functions is known as the Whittaker-Shannon sampling theorem.

Consider a function that is sampled by a Cartesian lattice of samples to give

$$\begin{aligned} f_S(x,y) &= \frac{1}{|a|}comb\left(\frac{x}{a}\right)\frac{1}{|b|}comb\left(\frac{y}{b}\right)f(x,y) \\ &= \frac{1}{|ab|}comb\left(\frac{x}{a},\frac{y}{b}\right)f(x,y) \end{aligned} \tag{3.29}$$

The sampled function is an array of delta functions spaced in the x direction by intervals a and spaced in the y direction by intervals b. They are weighted (areawise) by the values of the original function. Using the convolution theorem, the spectrum of the sampled function is

$$F_S(\xi,\eta) = comb(a\xi, b\eta) ** F(\xi,\eta) \tag{3.30}$$

The convolution of delta functions (2.14) along with the shift invariant property (2.13) gives the spectrum of (3.30) that is simply the spectrum of the original function located at each of the delta functions:

$$F_S(\xi,\eta) = \frac{1}{|ab|}\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} F\left(\xi - \frac{\eta}{a}, \eta - \frac{m}{b}\right) \tag{3.31}$$

Note that the sampled spectrum is a replication of the original function located at intervals of $1/a$ and $1/b$, as shown in Figure 3.8.

The scaling theorem relates the sample distances a and b to the spectrum spacing of (3.31). Note that if a and b are small, then the spacing in the frequency domain is large and vice versa. Therefore, a and b must be small enough to ensure that the frequency domain spacing does not allow overlap of the replicated spectrums. Overlap corresponds to information corruption that cannot be recovered. The requirements on a and b can be determined in the frequency domain such that

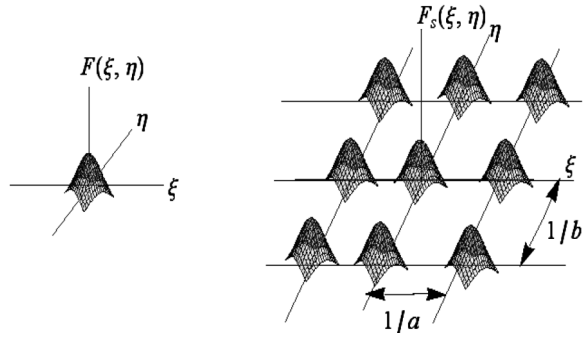


Figure 3.8 Spectra of (a) the original function and (b) the sampled function.

the frequency spacing is twice the function's spectrum widths. Let $\xi_{\max}$ and $\eta_{\max}$ be the band-limited widths of the original function's spectrum. Therefore,

$$a \le \frac{1}{2\xi_{\max}} \text{ and } b \le \frac{1}{2\eta_{\max}} \tag{3.32}$$

These are the maximal spacings of the samples for exact recovery of the original function. The sampled spectrum must be filtered to recover the original spectrum. A number of filters accomplish this task; however, perhaps the simplest function is

$$H(\xi,\eta) = rect\left(\frac{\xi}{2\xi_{\max}}, \frac{\eta}{2\eta_{\max}}\right) \tag{3.33}$$

When this transfer function is multiplied by the sampled function spectrum, the original function spectrum is obtained:

$$F(\xi,\eta) = F_S(\xi,\eta) rect\left(\frac{\xi}{2\xi_{\max}}, \frac{\eta}{2\eta_{\max}}\right) \tag{3.34}$$

This recovery of the original signal can also be seen in the spatial domain

$$f(x,y) = \left[f(x,y)\frac{1}{|ab|} comb\left(\frac{x}{a}\right) comb\left(\frac{y}{b}\right)\right] ** 4\xi_{\max}\eta_{\max} sinc(2\xi_{\max}x, 2\eta_{\max}y) \tag{3.35}$$

The sampled function $f_S(x,y)$ can be written as

$$\begin{aligned} f_S(x,y) &= f(x,y)\left(\frac{1}{|ab|}\right) comb\left(\frac{x}{a}\right) comb\left(\frac{y}{b}\right) \\ &= \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} f(na,mb)\delta(x-na, y-mb) \end{aligned} \tag{3.36}$$

Note that the sampled function is a weighted collection of delta functions and that the convolution of the *sinc* function with these delta functions, as given in (3.35), results in weighted *sinc* functions located at the delta functions. The reconstructed original signal is

$$f(x,y) = 4\xi_{\max}\eta_{\max} \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} f(na,mb) sinc[2\xi_{\max}(x-na), 2\eta_{\max}(y-mb)]) \tag{3.37}$$

If the sampling intervals are taken to be the maximum allowable spacing within the limits of (3.32), then (3.37) becomes

$$\begin{aligned} f(x,y) = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} & f\left(\frac{n}{2\xi_{\max}}, \frac{m}{2\eta_{\max}}\right) \\ & \times sinc\left[2\xi_{\max}\left(x-\frac{n}{2\xi_{\max}}\right), 2\eta_{\max}\left(y-\frac{m}{2\eta_{\max}}\right)\right] \end{aligned} \tag{3.38}$$

This identity is known as the Whittaker-Shannon [6] sampling theorem. It states that exact replication of a band-limited image function can be performed simply by applying a *sinc* interpolation function to the sampled image. Goodman [6] points out that this is not the only sampling theorem. It applies to the Cartesian sampling array and the rectangular transfer function. These choices could be altered to give a number of sampling theorems. In fact, the Whittaker-Shannon sampling theorem is widely applied to imaging systems, but does not work that well in sensor design, for reasons that are covered in Section 3.9.

3.8 Spatial Sampling and Resolution

A sensor's resolution characteristics depend on a number of variables, but can be discussed using two general parameters: the *system transfer function* and the *sampling function*. The system transfer function [16] includes optical degradation (diffraction and aberrations), detector size and shape, electronics and processing, display, and the human visual transfer function. The individual transfer functions can be cascaded to obtain a system transfer function (i.e., system MTF) that can be multiplied by an input spectrum to give the output spectrum. The sampling function (for sampled systems) corresponds to the discrete nature of spatial sampling. The sampling function has no transfer function but can definitely affect the spatial resolution of the system [4, 10].

Consider the sampled imaging system shown in Figure 3.9. A target is imaged onto a discrete detector array (a staring array)by imaging optics. We assume that the detector responses are uniform over the detector surfaces, the output light is the integrated light on the detectors, and the imaged surface is a large array of detectors. The image samples, or voltage, provided by the output of the detector array can be written as

$$i(x,y) = K\left[o\left(\frac{x}{c},\frac{y}{c}\right) ** rect\left(\frac{x}{p},\frac{y}{p}\right) ** h(x,y)\right]\frac{1}{p^2} comb\left(\frac{x}{p},\frac{y}{p}\right) \tag{3.39}$$

where K is a radiometry constant to describe the signal throughput of the system, p is the distance between pixel centers (pixel pitch), c is an image magnification constant equal to the focal length over the object range f/R and $h(x, y)$ is the optical impulse response or psf. The GSD acronym is known as the ground sampled distance (derived for ground-looking sensors on airborne platforms). The GSD is the projection of the pixel pitch onto the ground or target of interest at the object range. This is the absolute limit to the spatial resolution of the system. Many performance parameters such as the National Imagery Interpretation Rating System (NIIRS) [18] have been developed based on the GSD. Systems can have less resolution than the GSD because of the transfer function and corresponding psf. Projection of a psf onto the ground in the same manner as the sample function gives a feel for the associated resolution. If the projected psf is smaller than twice the GSD, then the resolution is limited by the sampling function. If the psf is larger than twice the GSD, then the resolution is limited by the transfer function. For example, *ground resolved distance* (GRD) is a photo reconnaissance parameter that corresponds to the mini-

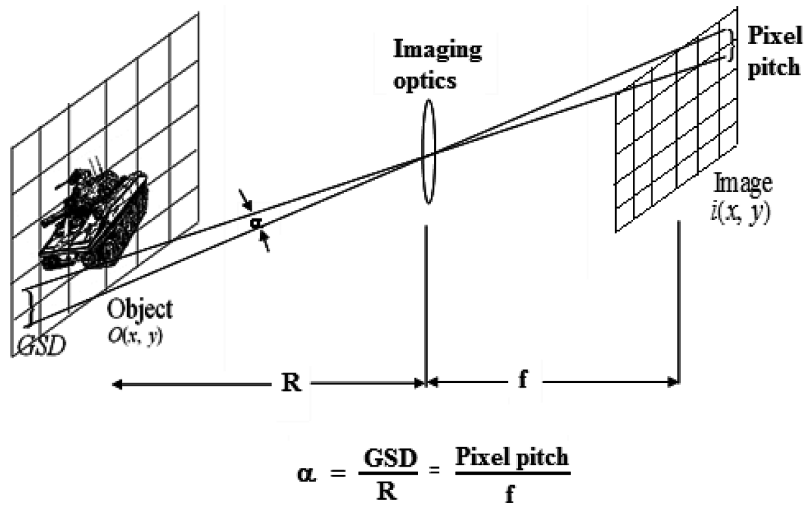


Figure 3.9 Sampled imaging system.

mum test target size that can be resolved on the ground by an experienced photo-interpreter. This parameter can be equivalent to the GSD if the psf is smaller than twice the pixel pitch. Optimal design gives a psf at twice the pixel pitch.

The sample image in the frequency domain can be expressed as

$$I(\xi,\eta) = Kc^2p^2\{O(c\xi,p\eta)sinc(p\xi,p\eta)H(\xi,\eta)\} **comb(p\xi,p\eta) \tag{3.40}$$

where $H(.,.)$ is the incoherent transfer function of the optics (i.e., the transform of the psf). The width of the *sinc* function is twice the width (distance between samples) of the *comb* function. From the Whittaker-Shannon sampling theorem, any function to be convolved with the sampling spectrum has to be less than half the sampling spacing in the frequency domain to retain all spatial frequency information. Therefore, the product *sinc H* must have a width less than half the *comb* spacing to retain information with no overlap of the convolved functions. If not, this overlap causes high-frequency information to mix with low-frequency information resulting in low-frequency artifacts. This effect is known as *aliasing*. Noting that the *sinc* function does not provide the necessary spectral limit to prevent aliasing, the transfer function must be designed to provide this requirement. In fact, most detectors do not completely fill the array and are smaller in width than the pixel pitch causing the size of the *rect* function to be smaller than that of (3.39). The smaller detector corresponds to a wider *sinc* in the frequency domain. Another solution, but not a practical one, is for the sampling system to view only low-frequency images with no significant high-frequency characteristics.

Finally, the spatial frequency of the object space can be given in a number of parameters. In the most basic form, irradiance can be written as a function of space, where x and y are in meters, so the spatial frequency parameters are in cycles per meter. A more common approach is to give x and y in radians or milliradians. That is, x and y in meter divided by the range from the sensor to the object yields a new x and y in radians (for small angles). To get milliradians, radians are multiplied by 1,000. The object irradiance as a function of x and y in milliradians gives ξ and

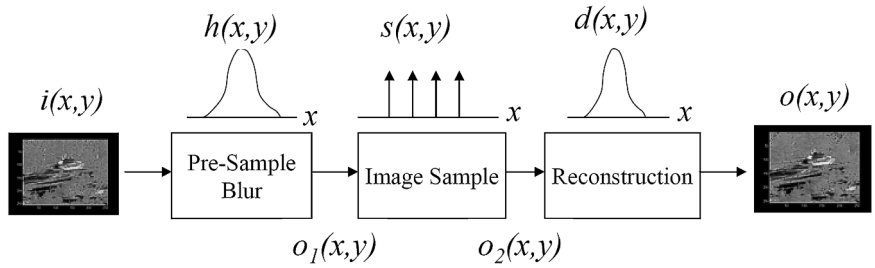


Figure 3.10 Three-step imaging process.

η in cycles per milliradian. Because these spatial frequencies are range independent, they are easier to apply in sensor performance calculations.

3.9 Sampled Imaging Systems

The previous few sections have described sampling using rough, but useful approaches. A more thorough treatment of sampling is supplied by Vollmerhausen [19], whose entire book is dedicated to a complete treatment of sampling. Here, we provide the basis for why sampled imaging systems can be considered linear, but not shift invariant.

The introduction of infrared focal planes over the past 10 years has resulted in staring array imagers, and for staring array imagers the application of LSI system analysis is of limited usefulness. These staring arrays have large detectors (15 to 20 μm or greater) with high fill factors. Because the sample spacing is at least as large as the detectors, these systems can be undersampled, which occurs if the optics' psf is smaller than the detector sample spacing.

Previously, we described the process of imaging for a continuous or well-sampled imager. In this case, the input scene is convolved with the imager psf (i.e., the impulse response of the system). With sampled imaging systems, the process is different. As an image traverses through a sampled imaging system, the image undergoes a three-step process. Figure 3.10 shows this process as a presample blur, a sampling action, and a postsample blur (reconstruction).

The image is blurred by the optics, the detector angular subtense, the spatial integration scan (if needed) and any other effects appropriate to presampling. This presample blur, $h(x,y)$, is convolved with the input scene:

$$o_1(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x-\alpha, y-\beta)h(\alpha,\beta)\, d\alpha\, d\beta = i(x,y) ** h(x,y) \tag{3.41}$$

where $o_1(x,y)$ is the presampled blur image or the output of the presample blur process. The sampling process can be modeled with the multiplication of the presample blur image with the sampling function. Here the sampling function is modeled in terms of the 1D *comb* function defined in Chapter 2:

$$comb\left(\frac{x}{b}\right) = |b| \sum_{n=-\infty}^{\infty} \delta(x - nb) \tag{3.42}$$

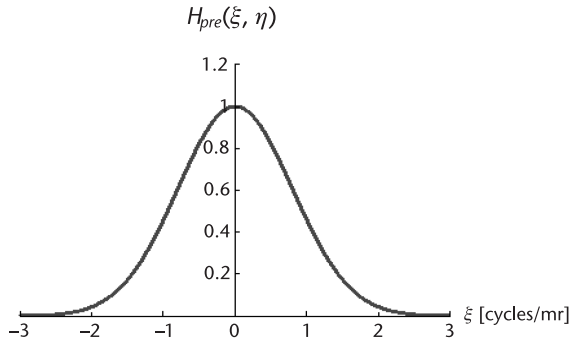


Figure 3.11 Presample blur transfer function.

The 2D *comb* function given in Table 2.4 is the product of two 1D *comb* functions. The output of the sampling process is written as the product of the presampled blurred image and the 2D *comb* function (note that a and b is the distance in milliradians or millimeters between samples):

$$\begin{aligned} o_2(x,y) &= o_1(x,y)\frac{1}{ab}comb\left(\frac{x}{a},\frac{y}{b}\right) \\ &= [i(x,y)**h(x,y)]\frac{1}{ab}comb\left(\frac{x}{a},\frac{y}{b}\right) \end{aligned} \tag{3.43}$$

The signal $o_2(x,y)$ is a set of discrete values that represent the presample blurred image at discrete locations. This output can be thought of as a weighted "bed of nails" that is difficult to interpret unless the "image" is reconstructed. The display and the eye, if applied properly, reconstruct the image to a function that is interpretable. This reconstruction is modeled as the convolution of the display and eye blur (and any other spatial postsample blur) and the output of the sampling process:

$$\begin{aligned} o(x,y) &= o_2(x,y)**d(x,y) \\ &= \left\{[i(x,y)**h(x,y)]\frac{1}{ab}comb\left(\frac{x}{a},\frac{y}{b}\right)\right\}**d(x,y) \end{aligned} \tag{3.44}$$

We have simplified (3.44) with the aggregate presample blur effects and the aggregate postsample reconstruction effects. The frequency analysis of the three-step process shown in Figure 3.10 can be presented simply by taking the Fourier transform of each process step. Consider the first step in the process, the presample blur. The transform of the convolution in space is equivalent to a product in spatial frequency

$$O_1(\xi,\eta) = I(\xi,\eta)H_{pre}(\xi,\eta) \tag{3.45}$$

were ξ and η are the horizontal and vertical spatial frequencies. If x and y are in milliradians, then the spatial frequencies are in cycles per milliradian. The term $H_{pre}(\xi, \eta)$ represents the Fourier transform of the presample blur spot. Note that the output spectrum can be normalized to the input spectrum in (3.45) so that $H_{pre}(\xi, \eta)$ is a transfer function that follows linear systems principles. Consider the

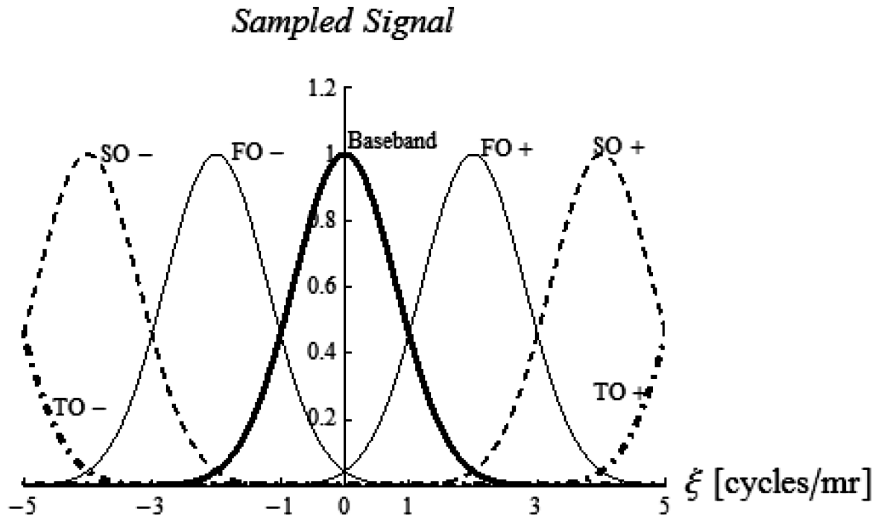


Figure 3.12 Sampling output.

presample blur spectrum (i.e., the pre sample blur transfer function) given in Figure 3.11. Note that this is the image spectrum on the output of the blur that would occur if an impulse were input to the system.

Next, we address the sampling process. The Fourier transform of (3.43) gives

$$O_2(\xi,\eta) = [I(\xi,\eta)H_{pre}(\xi,\eta)] ** comb(a\xi, b\eta) \tag{3.46}$$

If an impulse were input to the system, the response would be

$$O_2(\xi,\eta) = H_{pre}(\xi,\eta) ** comb(a\xi, b\eta) \tag{3.47}$$

which is a replication of the presample blur with sample spacing $1/a$ and $1/b$. Consider the case shown in Figure 3.12. The sampled spectrum shown corresponds to a Gaussian blur of 0.5 milliradian radius (to the 0.043 cutoff) and a sample spacing of 0.5 milliradian.

Note that the reproduction in frequency of the presample blur is 2 cycle/mr, the reciprocal of the sample rate. The so-called Nyquist rate of the sensor (really, it is the sensor half-sample rate) is at 1 cycle/mr. Any frequency f from the presample blur baseband that is greater than the half-sample rate is also present as a "mirror" signal, or classical aliasing, under the half-sample rate by the first-order reproduction. The amount of classical aliasing is easily computed as the area of this mirrored signal. However, this is not the aliased signal seen on the output of the display because the display transfer has not been applied to the signal. The higher order replications of the baseband are real frequency components. The FO+, SO+, FO–, and SO– are the first- and second-order replications at the positive and negative frequencies, respectively. The sample o_2 is tiny infinitesimal points weighted with the presample blurred image values. These points have spectra that extend in the frequency domain to very high frequencies. The higher order replications are typically filtered with a reconstruction function involving the display and the eye. There is no practical way to implement the perfect reconstruction filter, whereas the perfect

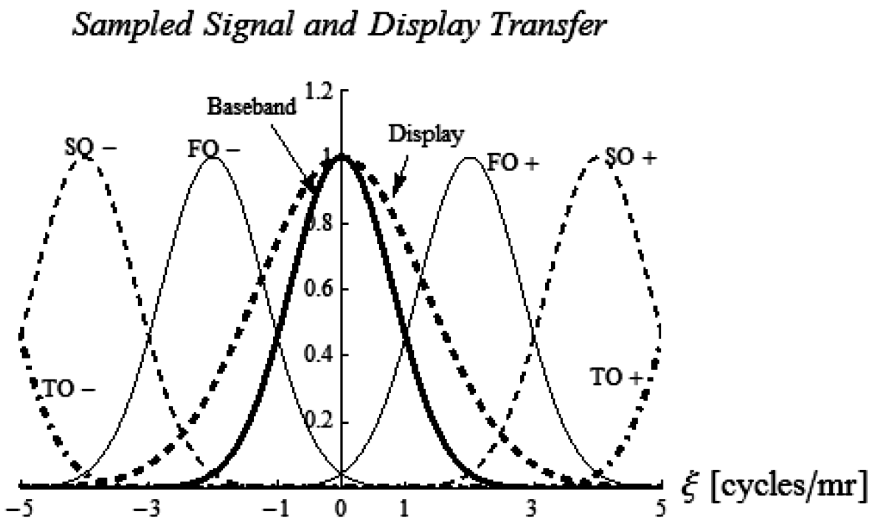


Figure 3.13 Sampled signal and display transfer.

rectangular filter would eliminate these higher order replications and result only in the classical aliased signal. The reconstruction filter usually degrades the baseband and allows some of the signal of the higher order terms through to the observer.

The final step in the process corresponds to the reconstruction of the sampled information. This is accomplished simply by blurring the infinitesimal points to where the function looks nearly like that of the continuous input imagery. Design guidelines are given for this reconstruction by Schade, Legault, and Sequin. The blur is convolved in space, so it is multiplied in frequency:

$$O(\xi,\eta) = [H_{pre}(\xi,\eta) * * comb(a\xi, b\eta)]D(\xi,\eta) \tag{3.48}$$

where this output corresponds to a point-source input. Note that the postsampling transfer function is multiplied by the sampled spectrum to give the output of the whole system. Consider the sampled spectrum and the dashed display transfer function shown in Figure 3.13. The postsampling transfer function (shown in the graph as the display) passes part of the first-order replications. However, this display degrades the baseband signal relatively little. The trade-off here is baseband resolution versus spurious response content.

Given that the signals traverse through the postsample blur's transfer function, the output is shown in Figure 3.14. There is classical aliasing on the output of the system, but a large signal corresponds to higher order replication signals that passed through the display. Aliasing and the higher order signals are collectively the spurious response of the sensor. These signals were not present on the input imagery, but are artifacts on the output imagery. Without sampling, these spurious signals would not be present. The higher order replicated signals are combined at each spatial frequency in terms of a vector sum, so it is convenient to represent the magnitude of the spurious signals as the root-sum-squared of the spurious orders that make it through the reconstruction process.

Three aggregate quantities have proven useful in describing the spurious response of a sampled imaging system: total integrated spurious response as defined

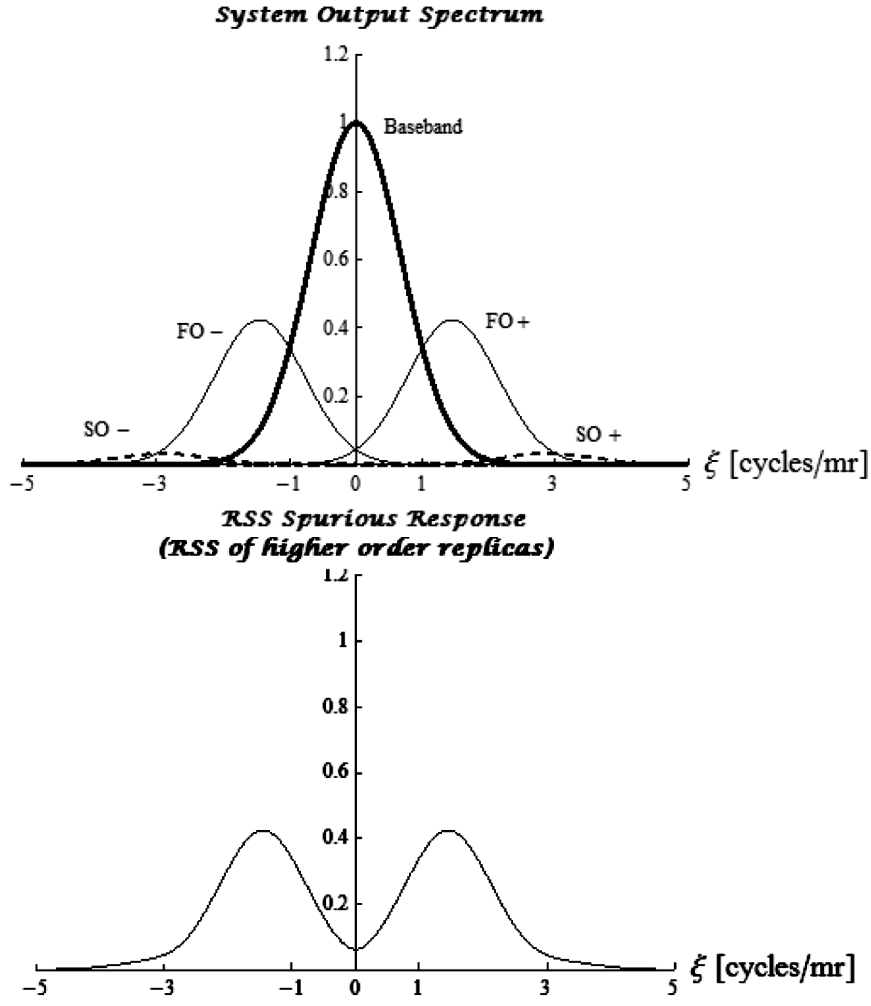


Figure 3.14 System output, RSS, and RSS spurious response spectra.

by (3.49a), in-band spurious response as defined by (3.49b), and out-of-band spurious response as defined by (3.49c):

$$SR = \frac{\int_{-\infty}^{\infty} (\text{spurious response})\, d\omega}{\int_{-\infty}^{\infty} (\text{baseband signal})\, d\omega} \tag{3.49a}$$

$$SR_{in\text{-}band} = \frac{\int_{-v/2}^{v/2} (\text{spurious response})\, d\omega}{\int_{-\infty}^{\infty} (\text{baseband signal})\, d\omega} \tag{3.49b}$$

$$SR_{out\text{-}of\text{-}band} = SR - SR_{in\text{-}band} \tag{3.49c}$$

The spurious responses of the higher order replications could be constructive or destructive in nature depending on the phase and frequency content of the spurious signals (that is the position; this is the primary reason why these systems are not shift invariant). The construction or destruction of bar patterns depends on the phase of the bar pattern with respect to the staring array sampling grid. The combination in magnitude is identical to that of a vector sum. The magnitude described in the previous equations is the quadrature sum of the signals. This integrated RSS spurious response value was normalized by the integral of the baseband area. This ratio was the spurious response ratio. These values describe in a reasonable way how much corruption is inherent in the image. In-band spurious response can be thought of as shifting object points, lines, and edges; in-band aliasing can also modify the apparent width of an object or make the object disappear. That is, a fence post imaged by an under sampled sensor can be seen as fatter, skinnier, or slightly misplaced. In-band aliasing generates localized disturbances or artifacts in the image; in-band aliasing represents corruption of scene elements. Out-of-band spurious response is a pattern or artifact that is caused by the display. For example, raster on a CRT display or the sharp pixel edges on a low-resolution flat panel display represent out-of-band spurious response.

3.10 Guide to the References

Books that discuss linear systems with an emphasis on the time domain include those by Strum [17], Oppenheim [13], Lathi [9], and Papoulis [14]. Bracewell [1, 2], Gaskill [5], Goodman [6], and Papoulis [15] extend linear systems theory to two dimensions.

3.11 Exercises

3.1 The Fourier transform can be considered as a system that maps a function into its transform.

a) Is this system linear?

b) Is this system shift invariant?

3.2 A system is specified by the operator:

$$S\{f(x)\} = 0.5\int_{-\infty}^{x} f(\alpha)\,d\alpha$$

a) Is the system linear?

b) Is the system shift invariant?

c) Sketch the output for $f(x) = rect\left(\frac{x-2}{2}\right)$

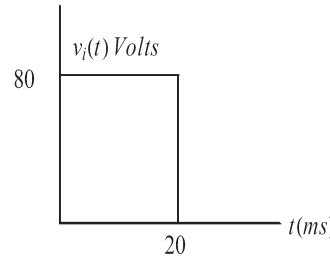
d) Sketch the output for $f(x) = rect\left(\frac{x+2}{2}\right)$

3.3 A system is specified by the operator:

$$S\{f(x)\} = [f(x)]^3 + [f(x)]^2 + 3[f(x)]$$

a) Is the system linear?
b) Is the system shift invariant?

3.4 Use the convolution integral to determine the output of a circuit if the voltage input is as shown below and the circuit impulse response is $h(t) = 10e^{-3t}$, $t \geq 0$.



3.5 An LSI system has an impulse response

$$h(x) = 4sinc^2(2x) - 2sinc^2(x)$$

The input to the system is $f(x) = 1 + 2\ cos\ (2\pi\xi_0 x)$. Sketch the output for
(a) $\xi_0 = 0.5$
(b) $\xi_0 = 1.0$
(c) $\xi_0 = 2.5$

3.6 The image irradiance formed by an incoherent optical system is given by

$$I(x,y) = |h_{coh}(x,y)|^2\ **|\ f(x,y)|^2 = h_{incoh}(x,y)**|\ f(x,y)|^2$$

where $h_{coh}(x,y)$ is the coherent impulse response, $h_{incoh}(x,y)$ is the incoherent impulse response (or psf), and $f(x,y)$ is the input image field amplitude. That is, the system is linear in irradiance. The output amplitude formed by a coherent system is

$$g(x,y) = h_{coh}(x,y)\ **\ f(x,y)$$

where the corresponding irradiance is

$$I(x,y) = |g(x,y)|^2$$

The system is linear with field amplitude. For

$$f(x,y) = rect\left(\frac{x}{\Delta x}\right)$$

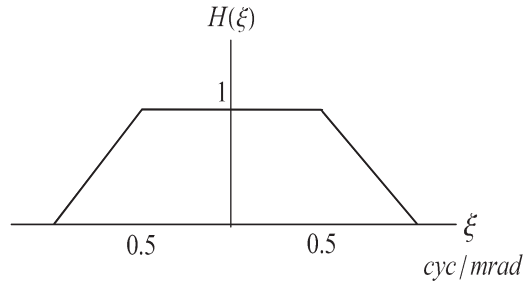
$$h(x,y) = rect\left(\frac{x}{\Delta x}\right) - rect\left(\frac{x-\Delta x}{\Delta x}\right)$$

(a) Find the system output irradiance formed by the coherent system.
(b) Find the system output irradiance formed by the incoherent system.

3.7 A system with the transfer function shown below is stimulated with input signals described by $f(x) = 1 + cos\ (2\pi\, \xi\, x)$ where

(a) $\xi = 0.5$
(b) $\xi = 0.75$
(c) $\xi = 1.0$

Determine and sketch the output signal for each case.



References

[1] Bracewell, R., *Two-Dimensional Imaging*, Upper Saddle River, NJ: Prentice Hall. 1995.

[2] Bracewell, R., *The Fourier Transform and Its Applications*, New York City, NY: McGraw-Hill. 2000.

[3] Cogdell, J. R., *Foundations of Electrical Engineering*, Englewood Cliffs, NJ: Prentice-Hall, 1990.

[4] Lloyd, J. M., "Fundamentals of Electro-Optical Imaging Systems Analysis," in *Infrared and Electro-Optical Systems* Handbook, p. 45, M. C. Dudzik (ed), Bellingham, WA: SPIE Press, 1993.

[5] Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics*, John Wiley& Sons, 1978.

[6] Goodman, J. W., *Introduction to Fourier Optics*, San Fransisco, CA: McGraw-Hill, 1968.

[7] Hayt, W., and J. Kemmerly, *Engineering Circuit Analysis*, NY: McGraw-Hill, 1978.

[8] Holst, G. C., *Electro-Optical Imaging System Performance*, SPIE, 2008.

[9] Lathi, B. P., *Linear Systems and Signals*, Oxford University Press, 2005.

[10] Lloyd, J., *Thermal Imaging Systems*, New York City, NY: Plenum Press, 1975.

[11] McDonough, R. N., and A. D, Whalen. *Detection of Signals in Noise, Second Edition*, San Diego, CA: Academic Press, 1995.

[12] Nilsson, J. W., *Electric Circuits*, Reading, MA: Addison-Wesley, 1989.

[13] Oppenheim, A., A. Willsky and I. Young, *Signals and Systems*, Prentice Hall, 1983.

[14] Papoulis, A., The *Fourier Integral and Its Applications*, McGraw-Hill, 1962.

[15] Papoulis, A., *Systems and Transforms with Applications in Optics*, Malabar, FL: Krieger Publishing, 1981.

[16] Ratches, J., "Static performance model for thermal imaging systems," *Optical Engineering*, 15(6):525–530, 1976.

[17] Strum, R. D., and D. E. Kirk. *Contemporary Linear Systems Using MATLAB*, Grove, CA: Brooks Cole, 2000.

[18] Imagery Interpretation Rating Scale. Air standard agreement. AID STD 101/11A, 1980.

[19] Vollmerhausen, R. H., and R. G. Driggers, *Analysis of Sampled Imaging Systems*, Bellingham, WA: SPIE Publications, 2000.

CHAPTER 4

Diffraction

The disturbance caused by a point source of light results in a plane wave at large distances. Consequently, if a plane wave is incident on an imaging system, then with an optical system that was perfectly designed using Snell's law and the law of reflection, it would be expected that all the rays would converge to a point. This expectation is not fulfilled. If the source is made arbitrarily small, even with a lens system that is perfectly designed from the point of view of Snell's law and the law of reflection, the image of the point source never gets smaller than a critical size that is predicted by diffraction theory.

The light distribution on the image plane from a point source is called the *point spread function* and is abbreviated *psf*. The psf is never smaller than a size predicted by diffraction theory and is a manifestation of the wave nature of the incident radiation. The psf size is proportional to the radiation wavelength, the optical focal length, and inversely proportional to the size of the optical system entrance pupil. The absolute value of the Fourier transform of the psf is termed the *optical modulation transfer function*. For an optical system where the psf has a size close to the minimum size predicted by diffraction theory, the Fourier transform of the psf is termed the *diffraction modulation transfer function*.

The optical or diffraction MTF represents the transfer of spatial frequencies through the optical system and is useful in analyzing imaging infrared and EO systems. This chapter is necessary for those who must deal with complex system apertures or those who simply desire a better understanding of this basic limitation to sensor performance.

Diffraction is a phenomenon that was discovered in the 1700s and studied by many great scientists of the time such as Young, Fresnel, Fraunhofer, Poisson, and Arago. Fresnel won the French national optics award for predicting a bright spot behind an illuminated disk that required the curving of light. This theory was considered beyond belief by Poisson, Laplace, Biot, and Gay-Lussac, so Arago (a judge at the competition) arranged for an immediate experimental demonstration in the basement of the conference hall. It was shown to be true and Fresnel's historical contribution became famous. In fact, this demonstration provided a significant shift in the Newtonian light paradigm of particle theory to a more wavelike characterization of light.

The essence of diffraction is that light spreads (or curves) around obstructions. This can be understood as the constructive interference of Huygens' wavelets. The spreading increases with longer light wavelengths and smaller obstructions. Constructive interference takes place not only at the expected location from the point of view of geometrical optics but in the vicinity of that point and makes it impossible to reconstruct an infinitesimal point source in the image plane of a sensor. The dimension over which constructive interference takes place is proportional to the wavelength of the light and would take place even if the optic is geometrically perfect, residing in vacuum.

Diffraction is one reason that large telescope apertures are desired. An optical system that is considered to be performing to its physical limit is called a diffraction-limited system. Large, ground-based astronomical telescopes are sometimes limited by optical disturbances in Earth's atmosphere. Prime locations for telescopes are high altitudes where atmospheric effects are reduced. Space-based systems such as the Hubble telescope are either diffraction limited or limited by the imperfections in the imaging paths of the optics (i.e., aberrations). The large apertures of these systems are designed to reduce the diffraction effects.

This chapter provides the fundamentals of diffraction analysis so that the student or analyst can determine diffraction effects for a particular sensor. The material here is limited to the basics needed to perform general sensor analysis. More specific information can be found in [1–18].

4.1 Electromagnetic Waves

In the context of electromagnetic radiation the term *waves* is usually short for *transverse electromagnetic (TEM) waves*. TEM simply means that the electric field is normal to the magnetic field and both fields are normal to the direction of wave propagation. We usually keep track of the electric field distribution because given it, and the direction of propagation, the magnetic field is known. For simplicity, the electric field distribution is described as a *complex amplitude*. The electric field irradiance is proportional to the squared absolute value of the complex amplitude.

The polarization of a wave indicates the direction of the electric field vector oscillation with respect to the direction of propagation. A linearly polarized wave is one in which this oscillation can be described on a single axis. For example, a wave propagating down the z axis of Figure 4.1 might be linearly polarized in the y direction. In this case, there would be no electric field oscillating in the x direction (i.e., only the magnetic field). There are a number of other polarization types, such as circularly polarized in the clockwise or counterclockwise directions, where the direction of the electric field vector changes as the wave propagates. Also, a random polarized wave would have an equal amount of electric field amplitude in all directions normal to the direction of propagation.

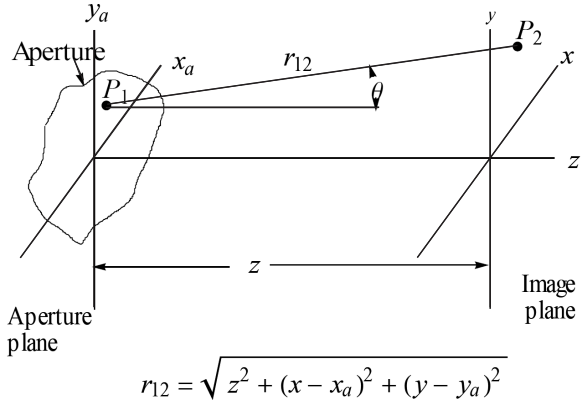


Figure 4.1 Diffraction geometry.

We can represent a monochromatic light wave using the scalar function

$$u(x,y,z,t) = a(x,y,z)cos(2\pi\upsilon t - \phi(x,y,z)) \tag{4.1}$$

where υ is the temporal frequency of the wave, $a(x,y,z)$ is the amplitude of the wave, and the argument of the cosine is the phase. Here $u(x,y,z,t)$ is termed the *optical disturbance* and is a function of position and time. The surfaces over which $\phi(x,y,z)$ is a constant are *wavefronts*. The wavefronts can take on planar, quadratic, spherical, cylindrical, and more complicated shapes. It follows from the Maxwell equations that in an isotropic medium, waves propagate in a direction perpendicular to the wavefront.

We use the phasor form, or complex amplitude, of the wave

$$u(x,y,z) = a(x,y,z)e^{j\phi(x,y,z)} \tag{4.2}$$

The temporal representation of the wave is retrieved from the phasor representation:

$$\begin{aligned} u(x,y,z,t) &= \mathrm{Re}\{u^*(x,y,z)e^{j2\pi\upsilon t}\} = \mathrm{Re}\{u^*(\boldsymbol{r})e^{j2\pi\upsilon t}\} \\ &= \frac{1}{2}\left(u^*(\boldsymbol{r})e^{j2\pi\upsilon t} + u(\boldsymbol{r})e^{-j2\pi\upsilon t}\right) \end{aligned} \tag{4.3}$$

The instantaneous irradiation $E(x,y,z,t)$ is

$$E(x, y, z, t) = u^2(x, y, z, t)$$

Detectors, including the human eye, cannot respond to the high temporal frequencies of visible or thermal radiation but instead respond to the time-averaged power of the radiation. The time-averaged irradiance $E(x,y,z)$ is

$$\begin{aligned} E(x,y,z) &= < E(x,y,z,t) > = < u^2(x,y,z,t) > \\ &= \frac{1}{4} < (u^*(\boldsymbol{r})e^{j2\pi\upsilon t} + u(\boldsymbol{r})e^{-j2\pi\upsilon t})^2 > \\ &= \frac{1}{4} < u^{*2}(\boldsymbol{r})e^{j4\pi\upsilon t} + u^2(\boldsymbol{r})e^{-j4\pi\upsilon t} + 2u(\boldsymbol{r})u^*(\boldsymbol{r}) > \\ &= \frac{1}{2}u(\boldsymbol{r})u^*(\boldsymbol{r}) = \frac{1}{2}\,|\,u(\boldsymbol{r})\,|^2 = |\,u_n(\boldsymbol{r})\,|^2 \end{aligned} \tag{4.4}$$

where we have introduced a *normalized optical disturbance* $u_n(\boldsymbol{r})$ related to the optical disturbance by

$$u_n(\boldsymbol{r}) \equiv \sqrt{2}u(\boldsymbol{r})$$

A plane wave traveling down the z axis of Figure 4.2 can be described as

$$u(x,y,z) = Ae^{jkz} \tag{4.5}$$

where k is the propagation vector, $k = 2\pi/\lambda$. The wavefronts are perpendicular to the z axis as shown in Figure 4.2. The wavefronts correspond to constant amplitudes

of the electric field and are separated by λ. Any phase change from z_1 to z_2 can be found from

$$\Delta\phi = (z_2 - z_1)k \quad \text{wave traveling in } z \text{ direction} \tag{4.6a}$$

The generalization of (4.5) to a wave traveling in the $\boldsymbol{k}$ direction is

$$u(x,y,z) = Ae^{j\boldsymbol{k}\boldsymbol{r}} = Ae^{jk(x\cos\theta_x + y\cos\theta_y + z\cos\theta_z)} = Ae^{jk(\gamma_x x + \gamma_y y + \gamma_z z)} \tag{4.7}$$

where (as indicated in Figure 4.3) θ_x, θ_y, and θ_z are the angles the $\boldsymbol{k}$ vector makes with the x,y,z axes; γ_x, γ, and γ_z are the direction cosines; and k is the magnitude of k.

The generalization of (4.6a) gives the phase difference between z_1 and z_2 for a wave traveling in an arbitrary direction:

$$\Delta\phi = k\gamma_z z_{12} \tag{4.6b}$$

The direction cosines satisfy the relationship

$$\gamma_x^2 + \gamma_y^2 + \gamma_z^2 = 1$$

This follows directly from the relationships

$$x^2 + y^2 + z^2 = r^2; \quad \cos\theta_x = \frac{x}{r}; \cos\theta_y = \frac{y}{r}; \cos\theta_z = \frac{z}{r} \tag{4.8}$$

Referring to Figure 4.1 and (4.7), the wave going from P_1 to P_2 is represented by

$$u(x,y,z) = Ae^{jk(\gamma_z z)}e^{jk(\gamma_x x + \gamma_y y)} \tag{4.9a}$$

Frequently we are interested in the optical disturbance in the x,y plane of Figure 4.1, and the distance between the aperture plane and the image plane contributes a constant phase factor that disappears when the irradiance is calculated using (4.4). In this case (4.9a) becomes

$$u(x,y) = A'e^{jk(\gamma_x x + \gamma_y y)} \quad \text{where } A' = Ae^{jk(\gamma_z z)} \tag{4.9b}$$

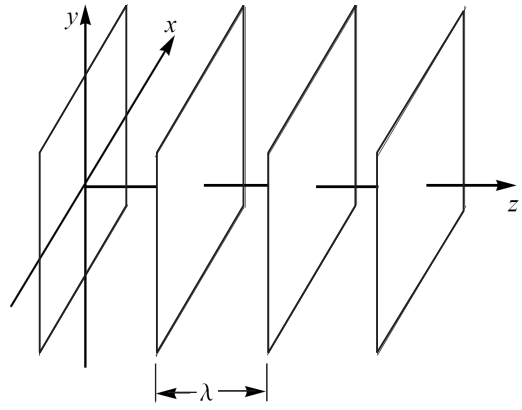


Figure 4.2 Plane wave propagation.

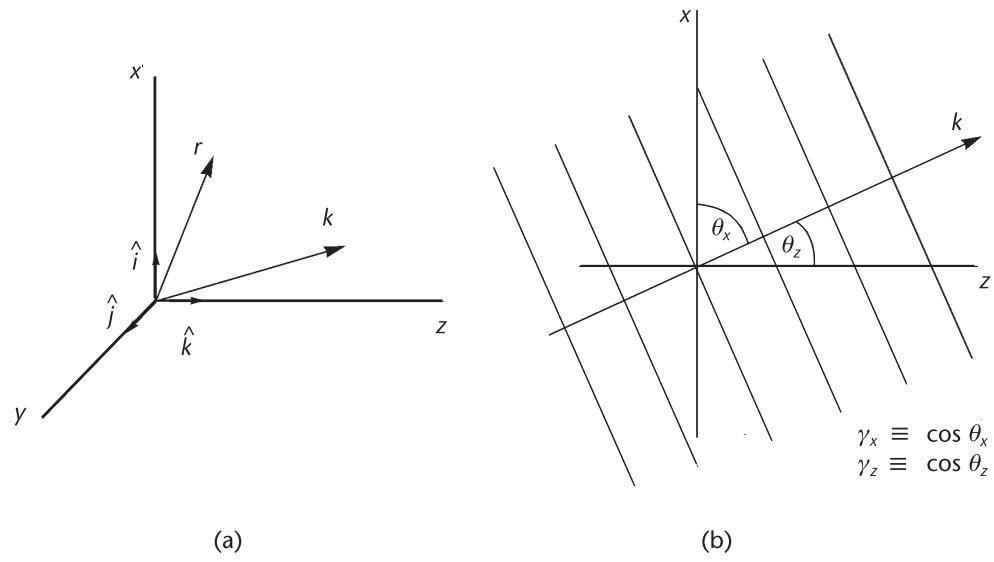


Figure 4.3 (a) Plane wave traveling in ***k*** direction. (b) Direction cosines.

4.2 Coherence

It is important to understand whether an imaging system is responding to coherent or incoherent light. Coherence defines the manner in which the optical transfer function is calculated. For a given aperture size, incoherent optical systems can resolve higher spatial frequencies than coherent systems, but there is a reduction in the modulation of higher spatial frequency signals compared with lower spatial frequency signals.

Coherence can be described as a fixed phase relationship of a field between two points in space or two points in time. The coherence [16] of a source determines the manner in which an irradiance distribution is calculated. At a system level, incoherent systems are linear in irradiance:

$$L\{a\,E_1(x,y)+b\,E_2(x,y)\}=a\,L\{E_1(x,y)\}+b\,L\{E_2(x,y)\} \tag{4.10}$$

A coherent system is linear in field distribution:

$$L\{a\,u_1(x,y)+b\,u_2(x,y)\}=a\,L\{u_1(x,y)\}+b\,L\{u_2(x,y)\} \tag{4.11}$$

In (4.11) the optical disturbances are normalized and from (4.4) the irradiance equals the square of the normalized optical disturbance. The irradiance from the sum of two fields for a coherent system is:

$$\begin{aligned} E_{total} &= (u_1(x,y)+u_2(x,y))((u_1(x,y)+u_2(x,y))^* \\ &= (u_1(x,y)u_1^*(x,y)+(u_2(x,y)u_2^*(x,y)+(u_1(x,y)u_2^*(x,y)+(u_1^*(x,y)u_2(x,y) \\ &= E_1(x,y)+E_2(x,y)+2\,\mathrm{Re}\{u_1(x,y)u_2^*(x,y)\} \end{aligned} \tag{4.12}$$

Equations (4.11) and (4.12) are illustrated in Figure 4.4. Here radiation from S gets to observation point P in the x,y plane via paths 1 and 2. Optical elements O_1 and O_2 cause the radiation to change direction by refraction, reflection, or diffraction. The last term in (4.12) describes the interference effects caused by the coherence of the source. Note that in an incoherent system, the interference terms are negligible. Coherence varies from no measurable phase relationship to a strong phase relationship. The term *partial coherence* is used to describe systems that are between coherent and incoherent. As explained below, the degree of coherence is determined by the source characteristics, spatial separation, and temporal separation.

The irradiance from the two sources when written in phasor notation, as in (4.12), suppresses the time dependence of the optical disturbance. Define a *correlation function* Γ_{12} at point P

$$\Gamma_{12}(\tau) \equiv < u_1(t) u_2^*(t+\tau) > \tag{4.13}$$

where $< \cdot >$ denotes a time average and τ is a time difference between u_1 and u_2 at P. A *normalized correlation function* γ_{12} is defined by

$$\gamma_{12}(\tau) \equiv \frac{\Gamma_{12}(\tau)}{\sqrt{E_1 E_2}} = \frac{< u_1(t) u_2^*(t+\tau) >}{\sqrt{E_1 E_2}} \tag{4.14}$$

Realize that in general $\Gamma_{12}(\tau)$ and $\gamma_{12}(\tau)$ are complex quantities and that $|\gamma_{12}(\tau)|$ is a dimensionless number between 0 and 1. When $\gamma_{12}(\tau)$ is one, then strong interference is observed at P. When $\gamma_{12}(\tau)$ is zero, then the two sources are incoherent and no interference pattern is observed at P. When $\gamma_{12}(\tau)$ is between zero and one, then there is partial interference at point P.

Using (4.14), (4.12) becomes

$$E_{total} = E_1(x,y) + E_2(x,y) + 2\sqrt{E_1(x,y) E_2(x,y)}\, \mathrm{Re}\{\gamma_{12}(\tau)\} \tag{4.15}$$

Equation (4.15) expresses (4.12) in terms of a normalized correlation function.

Let $E_{\max}$ and $E_{\min}$ denote the maximum and minimum irradiance of an interference pattern in the vicinity of P. The visibility V of the interference pattern at P is defined by

$$V \equiv \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}} \tag{4.16}$$

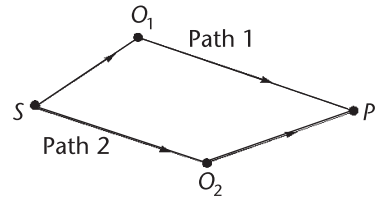


Figure 4.4 Radiation from two paths combined at *P*.

where V is a real number between 0 and 1. In (4.15), if $E_1 = E_2 = E_0$, then

$$\begin{aligned} E_{total} &= 2E_0(1 + \mathrm{Re}\{\gamma_{12}\}) \quad \text{so} \\ E_{\max} &= 2E_0(1 + |\gamma_{12}|) \\ E_{\min} &= 2E_0(1 - |\gamma_{12}|) \end{aligned} \tag{4.17}$$

and

$$V = \frac{4E_0|\gamma_{12}|}{4E_0} = |\gamma_{12}|$$

Equation (4.17) allows $|\gamma_{12}|$ to be calculated as a function of position from the visibility of interference fringes. Note that for incoherent sources, $|\gamma_{12}| = 0$ and the visibility of the fringes is zero, whereas for a coherent source, $|\gamma_{12}| = 1$ and the visibility of the fringes is one. See [13] for equations relating $|\gamma_{12}|$ to E_1 and E_2 for the case $E_1 \neq E_2$.

Two types of coherence are *spatial* and *temporal coherence*; however, both measure correlation. Spatial coherence is a measure of the phase correlation of the wave between two points in space along a wavefront. Temporal coherence is a measure of the correlation of the wave between two points in space in the direction of propagation, perpendicular to the wavefront. It will be shown later that (1) spatial coherence at a point P is determined by the angle subtended at P by the light source, and (2) temporal coherence is determined by the spectral width of the source.

A *monochromatic* source is one that emits at a single wavelength. Like the frictionless planes in mechanics, strictly speaking, monochromatic sources do not exist because even narrowband sources such as lasers emit over a narrow band of wavelengths. A monochromatic source would require an infinitely long wave. A *quasi-monochromatic* source is one that emits in a narrow band and this is a more proper way to describe narrowband radiation. For many applications, quasi-monochromatic sources can be modeled as monochromatic sources. Temporally coherent systems comprise monochromatic (single wavelength) sources. The system becomes partially temporally coherent as wavelengths are added and incoherent as a large number of wavelengths are added. Lasers exemplify partially temporal coherent sources because typically a laser is a narrow spectral bandwidth system.

A source can be modeled as one that emits sinusoidal wavetrains with lifetimes taken from some distribution. Four wavetrains are shown in Figure 4.5, as indicated by dashes, but typically there are many wavetrains. Let τ_0 denote the average lifetime for a

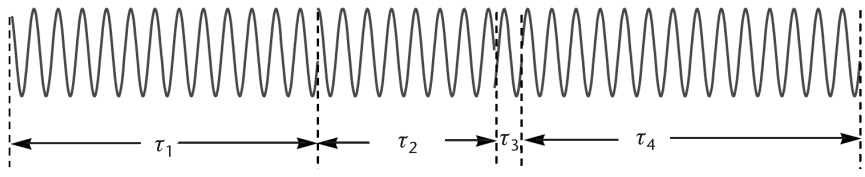


Figure 4.5 Wavetrains have different lifetimes or train lengths.

wavetrain. If a wave exists for a time τ_0 then there is an uncertainty Δf in the frequency of the wavetrain given by

$$\Delta f = \frac{1}{\tau_0} \tag{4.18}$$

A *temporal coherence length* l_t is the length of a coherent pulse:

$$l_t = c\,\tau_0 = \frac{c}{\Delta f} \tag{4.19}$$

Frequency, wavelength, and speed of propagation are related by $f = c/\lambda$. Differentiating this relationship and taking all differentials as being positive, we obtain

$$\Delta f = \frac{c}{\lambda^2}\Delta\lambda$$

Using this result, the temporal coherence length is determined from the source wavelength and the source line width:

$$l_t = \frac{\lambda^2}{\Delta\lambda} \tag{4.20}$$

Example 4.1
White light extends from approximately 0.4 to 0.7 μm with a center wavelength of 0.55 μm. Estimate the coherence length of white light.

The line width is 0.7 – 0.4 = 0.3 μm so $l_t = 0.55^2/0.3 \approx 1$ μm. The temporal coherence length of white light is about 1 μm or about twice the central wavelength. Other representative temporal coherence lengths are given by [13] and are exhibited in Table 4.1. Laser coherence lengths can be hundreds of meters, whereas the coherence length for a broadband source may be on the order of micrometers.

The need for spatial and temporal coherence concepts is illustrated in Figure 4.6. Radiation from a point source *S* goes through a double slit. We consider the radiation at a point *P*. Because slits *A* and *B* are equidistant from the point source, the radiation at slits *A* and *B* are spatially coherent. High visibility fringes will be seen at *P* if the path difference

$$\Delta \equiv S\,A\,P - S\,B\,P$$

is considerably less than the temporal coherence length; low visibility or no fringes will be seen at *P* if Δ is greater than the temporal coherence length. A point source

Table 4.1 Representative Coherence Lengths

Source	*Wavelength (μm)*	*Coherence Length*
White light	0.55	1 μm
Mercury light	0.546	1.2 cm
Krypton	0.606	78 cm
CO_2 laser	10.6	11 km

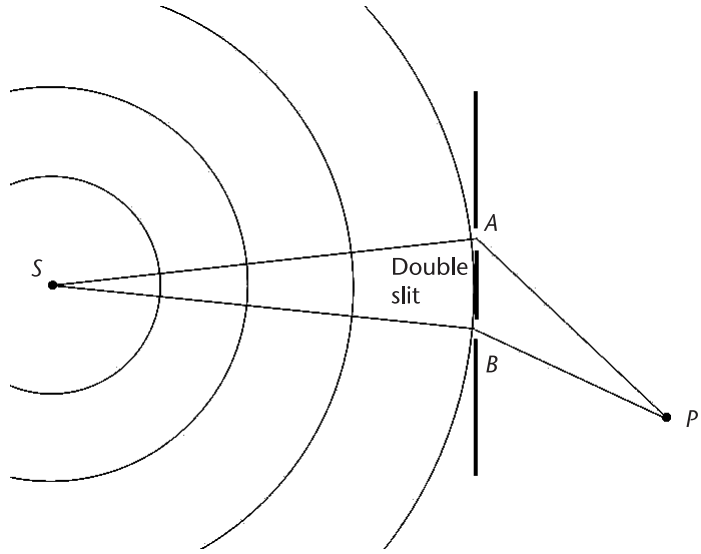


Figure 4.6 Temporal and spatial coherence concepts.

is an abstraction; all sources have finite extent. As the size of the source S grows, the radiation at slits A and B will no longer be spatially coherent and this influences the visibility of the fringes at P.

Referring to Figure 4.7, which illustrates two independent light sources S_1 and S_2 emitting radiation with wavelength λ separated from each other by a distance s and from the point P by a distance r. Let θ denote the angle subtended by S_1 and S_2 at P. Then the extent l_s of spatial coherence satisfies the relationship

$$l_s < \frac{\lambda}{\theta} \tag{4.21}$$

Reference Figure 4.8 to see why this is true. Monochromatic rays with wavelength λ from point sources A and B are incident on points P_1 and P_2. We wish to determine how the spatial coherence length l_s is related to the separation s of the point sources A and B. We can think of P_1 and P_2 as slits that form an interference pattern on the screen at point P. Here we assume the distance r is much greater than the separation

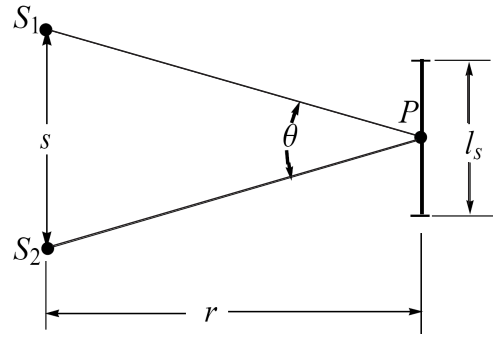


Figure 4.7 Source separation and spatial coherence length.

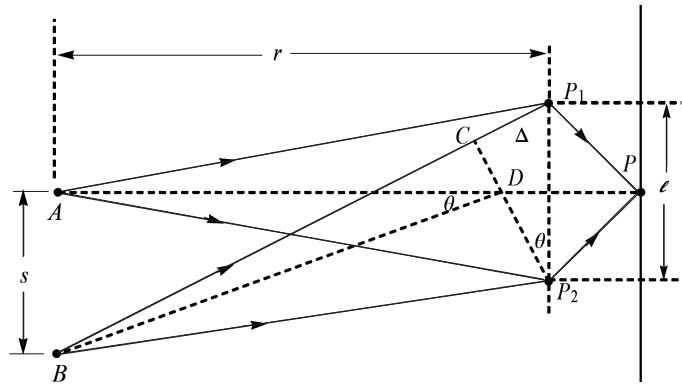


Figure 4.8 Spatial coherence length and source separation.

s of the sources or the separation ℓ of the slits. Because r is much greater than s, the rays BP_1 and BP_2 are nearly parallel, so P_2C is nearly perpendicular to BP_1 and Δ is the difference in path length between BP_1 and BP_2. Also BD is approximately perpendicular to $P_2\,C$ so the angle $C\,P_2\,P_1$ is very nearly the same as the angle $B\,D\,A$ and is denoted by θ in Figure 4.8. With s much less than r, the angle θ is near zero.

If s is zero, sources A and B coincide, $\Delta = 0$ and there will be a strong maximum at P. As s is increased, Δ grows to $\Delta = \lambda/2$. At that point the optical disturbance at point P from source A is cancelled by the optical disturbance from point B. At this separation of the two sources, the interference fringes on the screen would have little or no visibility so this distance ℓ is greater than l_s. From Figure 4.8:

$$sin\theta \approx \theta = \frac{\Delta}{\ell} \text{ so } \ell = \frac{\Delta}{\theta} = \frac{\lambda}{2\theta} > l_s$$

Here the calculation was done for two point sources. If there were a continuum of sources between points A and B then the factor 2 in the above equation goes away and we get (4.21).

Note that the imaging systems presented within the scope of this text can be considered to be incoherent systems. With objects scattered throughout a scene, in a typical scene, the sources subtend a large angle at the sensor and so (4.21) implies small spatial coherence lengths. Bandwidths are usually large compared to the central wavelength so (4.20) implies small temporal coherence lengths. The majority of practical imaging systems can be considered to be incoherent systems with the exception of those with extremely small sources or laser illuminated systems. The development of incoherent system theory, however, is better presented in terms of coherent systems. Note from Table 2.3 that the square magnitude relationship between field and irradiance [see (4.4)] can transform into a correlation in the frequency domain.

4.3 Fresnel and Fraunhofer Diffraction from an Aperture

We introduce diffraction first in terms of the coherent optical system response. The incoherent response function is then presented as the autocorrelation of

the coherent optical system response. Diffraction [4, 5] is a general phenomenon that applies not only to the field of light wave propagation, but also to sound and radio waves. The importance here is that diffraction must be understood to comprehend the limitations of optical systems. Diffraction is defined by Sommerfeld as "any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction." Diffraction theory allows one to calculate the light pattern caused by light as it passes through apertures or entrance pupils.

We consider here a very brief history of the current diffraction models, but will not present a mathematical derivation of the models used in this text. The interested reader should refer to [4, 5]. One such model starts with the application of Green's theorem to the Helmholtz wave equation to determine a complex field quantity in terms of its boundary values. The result was the integral theorem of Helmholtz and Kirchhoff. The Kirchhoff boundary conditions were then applied to this theorem to derive the first diffraction model, the Fresnel-Kirchhoff diffraction formula. This model provided very accurate results except where field quantities approached the obstruction boundaries. The inconsistencies were resolved by Sommerfeld with a shift in boundary phase terms and the current Rayleigh-Sommerfeld diffraction formula became the fundamental model for describing diffraction. Finally, the Huygens-Fresnel principle extended the diffraction theory to a superposition integral that generalized diffraction with the important linearity property. The reader is directed to Goodman [5] as a reference for these derivations. We begin here with the result of the Huygens-Fresnel principle.

Fresnel Diffraction

Consider the diffraction geometry shown in Figure 4.1. Using the Pythagorean theorem r_{12} is expressed in terms of the aperture plane coordinates x_a,y_a and the imaging plane coordinates x,y:

$$
\begin{aligned}
r_{12} &= \sqrt{z^2+(x-x_a)^2+(y-y_a)^2} = z\sqrt{1+\left(\frac{x-x_a}{z}\right)^2+\left(\frac{y-y_a}{z}\right)^2} \\
&= z\left(1+\frac{1}{2}\left\{\left(\frac{x-x_a}{z}\right)^2+\left(\frac{y-y_a}{z}\right)^2\right\}-\frac{1}{8}\left\{\left(\frac{x-x_a}{z}\right)^2+\left(\frac{y-y_a}{z}\right)^2\right\}^2+\cdots\right) \\
&\approx z\left(1+\frac{1}{2}\left\{\left(\frac{x-x_a}{z}\right)^2+\left(\frac{y-y_a}{z}\right)^2\right\}\right)
\end{aligned}
\tag{4.22}
$$

The algebraic manipulation in (4.22) was motivated by the observation that typically z $\gg$ max (x,y,x_a,y_a) so the quantity within { } is much less than one. In the second line the square root was represented by its Taylor series expansion. In the third line we utilized the observation that { } is much less than one.

In spherical coordinates the amplitude of a spherical wave moving away from the origin is represented by

$$u(x,y) = A\frac{e^{j(kr-\omega t)}}{r} \tag{4.23}$$

The amplitude is inversely proportional to r and this is a manifestation of the observation that the power per unit area flowing away from the origin is proportional to $|u|^2$ and the integral of this quantity must be constant over any sphere centered at the origin. The factor A enables the field strength u to have units appropriate to an optical disturbance.

Using Huygens' principle, with a plane wave incident normally on the aperture of Figure 4.1, the optical disturbance at P is the sum of spherical waves emanating from the aperture,

$$u(x,y) = Ae^{-j\omega t}\int\int_A \frac{e^{jkr_{12}}}{r_{12}}\,dx_a\,dy_a \tag{4.24}$$

and the integral is over the clear aperture A. We are interested in how $u(x,y)$ depends on the coordinates x,y in the image plane so we suppress the time dependence in (4.24). The function $1/r$ changes very slowly when r is large and so in the denominator of (4.24), r_{12} is well approximated by z, the first term in (4.22). If

$$\begin{aligned}&\frac{1}{8}kz\left\{\left(\frac{x-x_a}{z}\right)^2+\left(\frac{y-y_a}{z}\right)^2\right\}^2 \ll 1 \text{ or equivalently}\\ &z^3 \gg \frac{1}{8}k\left\{(x-x_a)^2+(y-y_a)^2\right\}^2\end{aligned} \tag{4.25}$$

for every point x_a,y_a in the aperture, then by suppressing the time dependence, (4.24) can be written:

$$u(x,y) = A\frac{e^{jkz}}{z}\cdot\int\int_A e^{j\frac{k}{2z}[(x-x_a)^2+(y-y_a)^2]}\,dx_a\,dy_a \tag{4.26a}$$

$$u(x,y) = A\frac{e^{jkz}}{z}\int\int_{-\infty}^{\infty}\tau(x_a,y_a)e^{j\frac{k}{2z}[(x-x_a)^2+(y-y_a)^2]}\,dx_a\,dy_a \tag{4.26b}$$

Equation (4.26a) describes how to calculate the amplitude $u(x,y)$ of the disturbance at P_2 in Figure 4.1 in the Fresnel approximation by integrating over the opening A in the aperture plane. In (4.26b) the integral is over the entire aperture plane and $\tau(x_a,y_a)$ is defined to be one where the aperture plane is clear and defined to be zero where the aperture plane is opaque.

There is an alternate way to express (4.26b). Expand the term in square brackets:

$$(x-x_a)^2+(y-y_a)^2 = x^2+y^2-2(x\,x_a+y\,y_a)^2+x_a^2+y_a^2$$

Using this result, (4.26b) can be written in the form

$$u(x,y) = A\frac{e^{jkz}}{z}e^{j\frac{k(x^2+y^2)}{2z}}\iint_{-\infty}^{\infty}\tau(x_a,y_a)\left\{e^{j\frac{k\left(x_a^2+y_a^2\right)}{2z}}\right\}e^{-j\frac{k}{z}[x\,x_a+y\,y_a]}\,dx_a\,dy_a$$
$$= A\frac{e^{jkz}}{z}e^{j\frac{k(x^2+y^2)}{2z}}\iint_{-\infty}^{\infty}\left\{\tau(x_a,y_a)e^{j\frac{k\left(x_a^2+y_a^2\right)}{2z}}\right\}e^{-j\frac{2\pi}{\lambda z}[x\,x_a+y\,y_a]}\,dx_a\,dy_a \tag{4.27}$$

Note that this is the Fourier transform of the quantity inside { }. The spatial frequencies $\xi = \frac{x}{\lambda z}$ and $\eta = \frac{y}{\lambda z}$ have the units of cycles per unit length.

Fraunhofer Diffraction

Suppose z satisfies (4.25) and also satisfies

$$\frac{k(x_a^2+y_a^2)}{2z} \ll 1 \;\Leftrightarrow\; z \gg \frac{k}{2}(x_a^2+y_a^2) \tag{4.28}$$

For representative aperture dimensions and field coordinates P_2, (4.28) is a more stringent requirement than (4.25). Assuming (4.25) and (4.28) are satisfied, (4.27) simplifies to:

$$u(x,y) = A\frac{e^{jkz}}{z}e^{j\frac{k(x^2+y^2)}{2z}}\iint_{-\infty}^{\infty}\tau(x_a,y_a)e^{-j\frac{2\pi}{\lambda z}[x\,x_a+y\,y_a]}\,dx_a\,dy_a \tag{4.29}$$

With the substitution

$$\xi = \frac{x}{\lambda z};\quad \eta = \frac{y}{\lambda z}\quad [\text{cycle/length}] \tag{4.30}$$

we see that $u(x,y)$ is proportional to the Fourier transform of the aperture function. The use of (4.29) will now be illustrated.

Example 4.2

Consider a rectangular aperture placed in the aperture plane shown in Figure 4.1. The aperture is irradiated with a monochromatic plane wave incident normally to the aperture plane. Assuming that Fraunhofer conditions are met, determine the irradiance distribution on the image plane.

Suppose the aperture has length a in the horizontal direction and width b in the vertical direction. Then

$$\tau(x_a,y_a) = rect\left(\frac{x_a}{a},\frac{y_a}{b}\right)$$

The double integral in (4.29) is the Fourier transform of $\tau(x_a,y_a)$ and is readily obtained from Tables 2.3 and 2.4:

$$\mathcal{FF}\left\{\tau\left(x_a,y_a\right)\right\} = a\,b\,sinc\left(a\,\xi,b\,\eta\right)\Big/\cdot\left\{\xi\to\frac{x}{\lambda z},\eta\to\frac{y}{\lambda z}\right\} = a\,b\,sinc\left(\frac{a\,x}{\lambda z},\frac{b\,y}{\lambda z}\right)$$

The field distribution in the image plane is given by (4.29):

$$u(x,y) = A\frac{e^{jkz}}{2z}e^{j\frac{k(x^2+y^2)}{z}}ab\,sinc\left(\frac{a\,x}{\lambda z},\frac{b\,y}{\lambda z}\right) \tag{4.31}$$

The irradiance distribution in the image plane is given by (4.4):

$$E(x,y) = \frac{A^2}{2}\frac{1}{z^2}a^2\,b^2\,sinc^2\left(\frac{a\,x}{\lambda z},\frac{b\,y}{\lambda z}\right) \tag{4.32}$$

Let us examine this equation to see if it makes sense. It seems reasonable that the irradiance is proportional to the square of the field strength parameter A and diminishes inversely as the square of the distance between the image and aperture plane. The field amplitude is reasonably proportional to the area of the aperture and so the irradiance distribution is proportional to the square of the aperture area. Examining the argument of the *sinc* function we note that the arguments are dimensionless as they should be. Note that the size of the image in the x direction is inversely proportional to the coefficient of x, which we observe equals $a/(\lambda z)$. Thus the larger a is, the smaller the diffraction pattern; the larger z is, the larger the diffraction pattern; and the larger λ is, the larger the diffraction pattern. Similar results are obtained in the y direction.

There is one feature that is not captured by (4.32). Suppose the monochromatic wavelength of the incident radiation is allowed to vary with the amplitude A independent of λ. Then, according to (4.32), the irradiance at the point $x = 0$, $y = 0$ is independent of λ. Experimentally it is found that under these conditions the irradiance at $x = 0$, $y = 0$ gets larger with smaller wavelength. The reason for the difference in the result displayed in (4.32) and experiment can be traced to (4.24). We used a simplified model of how the Huygens' wavelets at the aperture add to produce the field amplitude at point P_2. A more accurate formulation of (4.24) introduces a factor of $j\lambda$ in the denominator and an obliquity factor $q(\theta)$, which is one in the forward direction and zero in the backward direction:

$$u(x,y) = \frac{Ae^{-j\omega t}}{j\,\lambda}\int\int_A \frac{e^{jkr_{12}}}{r_{12}}q(\theta)\,dx_a\,dy_a \tag{4.33}$$

We are usually only interested in small forward angles so we set the obliquity factor $q(\theta)$ to one. Most of the time we are not interested in how the intensity for a given size aperture varies with λ. The factor of j in the denominator of (4.33) is due to the fact that the amplitude distribution in the image plane is related to the time derivative of the distribution in the aperture plane. The time derivative of a sinusoidal function is another sinusoid and in those cases the j in the denominator introduces an unimportant temporal phase change. The reader who desires a more comprehensive treatment of diffraction than that provided here is referred to [5].

4.4 Fraunhofer Diffraction from a Thin Lens

A topic of considerable importance is the effect of diffraction on optical systems. The traditional approach in the analysis of optical systems is to represent the lens as a transmittance function. Consider the lens shown in Figure 4.9, where the lens is positioned in the aperture plane.

The lens introduces a phase difference to rays that pass through the lens given by (see Goodman, [5])

$$t(x_a, y_a) = e^{jkn\Delta_0} e^{-j\frac{k}{2f}\left(x_a^2 + y_a^2\right)} \tag{4.34}$$

In this equation, $k = \dfrac{2\pi}{\lambda}$, n is the lens index of refraction, and f is the local length, with x_a,y_a being the coordinates of a point in the plane of the aperture and also within the lens. A limitation of (4.34) is that it applies to paraxial rays. Here we point out that (4.34) reasonably asserts that the maximum phase difference is the optical path length through the lens $n\,\Delta_0$ multiplied by k and the phase difference gets smaller for rays that traverse the lens father from the optical axis.

Substituting (4.34) into the Fresnel equation of (4.27) the optical disturbance at point P_2 with a lens filling the aperture is found:

$$u(x,y) = A\frac{e^{jkz}}{z} e^{j\frac{k(x^2+y^2)}{2z}} \int\int_{-\infty}^{\infty} \left\{ p(x_a, y_a) t(x_a, y_a) e^{j\frac{k(x_a^2+y_a^2)}{2z}} \right\} e^{-j\frac{2\pi}{\lambda z}[x\,x_a + y\,y_a]}\, dx_a\, dy_a$$

$$= Ae^{jkn\Delta_0} \frac{e^{jkz}}{z} e^{j\frac{k(x^2+y^2)}{2z}}$$

$$\int\int_{-\infty}^{\infty} \left\{ p(x_a, y_a) e^{-j\frac{k}{2f}(x_a^2+y_a^2)} e^{j\frac{k(x_a^2+y_a^2)}{2z}} \right\} e^{-j\frac{2\pi}{\lambda z}[x\,x_a + y\,y_a]}\, dx_a\, dy_a$$

Note that in the first line we replaced $\tau(x_a,y_a)$ with $p(x_a,y_a)t(x,y)$ to emphasize that the aperture is now the entrance pupil $p(x_a,y_a)$ and to explicitly note the phase change $t(x_a,y_a)$ introduced by the lens. Now suppose we choose the image plane to be a distance f behind the aperture plane, that is, $z = f$. Then the product of the last two factors within {} is one, so the equation simplifies to

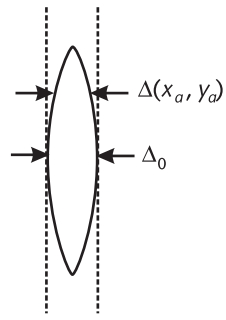


Figure 4.9 Lens transmittance.

$$u(x,y) = Ae^{jkn\Delta_0}\frac{e^{jkf}}{f}e^{j\frac{k(x^2+y^2)}{2f}}\int\int_{-\infty}^{\infty}p(x_a,y_a)e^{-j\frac{2\pi}{\lambda f}[x\,xa+y\,ya]}\,dx_a\,dy_a$$

$$= Ae^{jkn\Delta_0}\frac{e^{jkf}}{f}e^{j\frac{k(x^2+y^2)}{2f}}\mathcal{FF}\left\{p\left(x_a,y_a\right)\right\}/.\left\{\xi\to\frac{x}{\lambda f},\ \eta\to\frac{y}{\lambda f}\right\}$$

$$= Ae^{jkn\Delta_0}\frac{e^{jkf}}{f}e^{j\frac{k(x^2+y^2)}{2f}}P\left(\xi,\eta\right)/.\left\{\xi\to\frac{x}{\lambda f},\eta\to\frac{y}{\lambda f}\right\} \tag{4.35a}$$

Note that $u(x,y)$ is the optical disturbance associated with the psf. Also note that the pupil function $p(x_a,y_a)$ satisfies the condition

$$p(x_a,y_a) = \begin{cases}1 & \text{when the point } x_a,y_a \text{ is in the pupil}\\ 0 & \text{otherwise}\end{cases} \tag{4.35b}$$

In (4.35a) the integral is over the pupil function $p(x_a,y_a)$. If a transparency $t(x_a,y_a)$ multiplied the pupil function $p(x_a,y_a)$, which corresponds to having a transparency in the aperture plane the same size as the lens, then instead of $P(\xi,\eta)$ in the last line of (4.35a), we would have $P(\xi,\eta)$ ** $T(\xi,\eta)$. If $P(\xi,\eta)$ is approximately a delta function (corresponding to a large pupil), then at a distance f, the lens and the transparency approximately produce $T(\xi,\eta)$, the Fourier transform of the transparency. This is a fundamental result in the field of optical image processing.

4.5 Thin Lens Optical System Diffraction PSF

The system given in Figure 4.1, with a lens in the aperture and the image plane a distance f from the aperture plane, is the typical long-range imaging system. A distant point source on the optical axis of the lens produces a plane wave normally incident on the aperture plane and $u(x,y)$ is the optical disturbance at the image plane from that point source. As indicated by (4.4), the irradiance at the image plane is proportional to $u(x,y)u^*(x,y)$. The diffraction psf for an incoherent optical system is the irradiance normalized so that it is one at the origin (place where the optical axis intersects the image plane) and is easily calculated with the help of (4.35a):

$$psf(x,y) = \frac{u(x,y)u^*(x,y)}{u(0,0)u^*(0,0)} = \frac{\left|u(x,y)\right|^2}{\left|u(0,0)\right|^2}$$

$$= \frac{\left|\mathcal{FF}\left\{p(x,y)\right\}/.\left\{\xi\to\frac{x}{\lambda f},\ \eta\to\frac{y}{\lambda f}\right\}\right|^2}{\left|\mathcal{FF}\left\{p(x,y)\right\}/.\left\{\xi\to 0,\eta\to 0\right\}\right|^2} \tag{4.36}$$

We use this result to calculate the psf of a rectangular entrance pupil a units wide and b units high. For this case

$$p(x,y) = rect\left(\frac{x}{a},\frac{y}{b}\right)$$

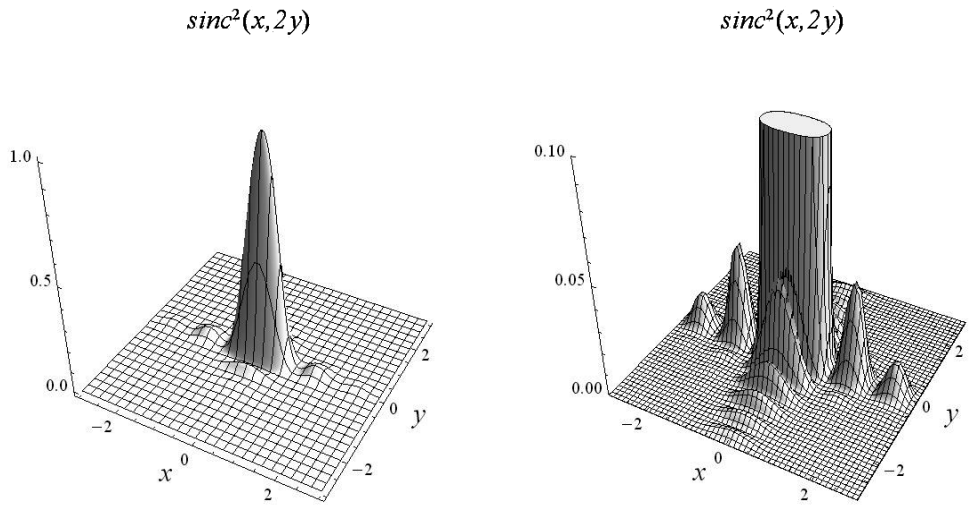


Figure 4.10 Point spread function for a rectangular aperture.

Then utilizing Tables 2.3 and 2.4 and (4.36), we compute the *psf*:

$$psf(x,y) = \frac{\left| ab\,sinc\left(a\xi, b\eta\right) /. \left\{ \xi \to \frac{x}{\lambda f}, \eta \to \frac{y}{\lambda f} \right\} \right|^2}{\left| ab\,sinc\left(a\xi, b\eta\right) /. \left\{ \xi \to 0, \eta \to 0 \right\} \right|^2} = sinc^2\left(\frac{ax}{\lambda f}, \frac{by}{\lambda f} \right)$$

Graphs of the rectangular aperture *psf* are shown in Figure 4.10 for the case where $\frac{a}{\lambda f} = 1, \frac{b}{\lambda f} = 2$.

We next calculate the psf of *a* circular aperture with diameter D. Using Figure 2.3 a circular pupil with diameter D has an entrance pupil given by

$$p(x,y) = cyl\left(\frac{r_a}{D} \right)$$

The radial coordinate for the pupil shape has an *a* subscript because it is in the aperture plane. Using Tables 2.3 and 2.4, we obtain

$$\begin{aligned} \mathcal{FF}\left\{ cyl\left(\frac{r_a}{D} \right) \right\} /. \rho \to \frac{r}{\lambda f} &= D^2 somb(D\rho) /. \rho \to \frac{r}{\lambda f} \\ &= D^2 \frac{2J_1\left(\pi D \rho\right)}{\pi D \rho} /. \rho \to \frac{r}{\lambda f} \\ &= D^2 \frac{2J_1\left(\frac{\pi D r}{\lambda f} \right)}{\frac{\pi D r}{\lambda f}} \end{aligned} \tag{4.37}$$

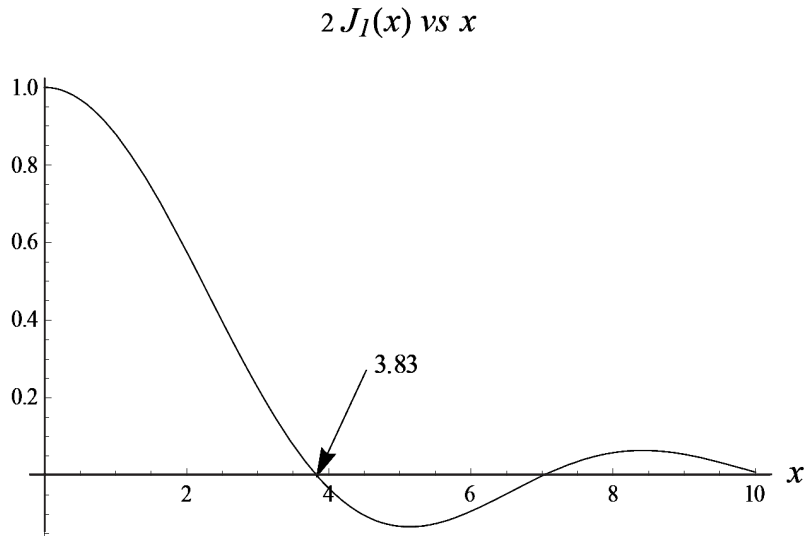


Figure 4.11 Amplitude of psf for a circular entrance pupil.

In (4.37), ρ is the spatial frequency in the radial direction in units of cycles per unit distance, r is the radial coordinate in the image plane, D is the diameter of the entrance pupil, λ is the wavelength of the radiation, and f is the focal length of the lens. The derivation of the point spread function exhibited here was considerably simplified by the use of Tables 2.3 and 2.4, which hides considerable mathematical detail. See Hecht and Zajac [7] for a derivation of (4.37) without the use of Tables 2.3 and 2.4. Note that

$$\lim_{x \to 0} \frac{2J_1(x)}{x} = 1$$

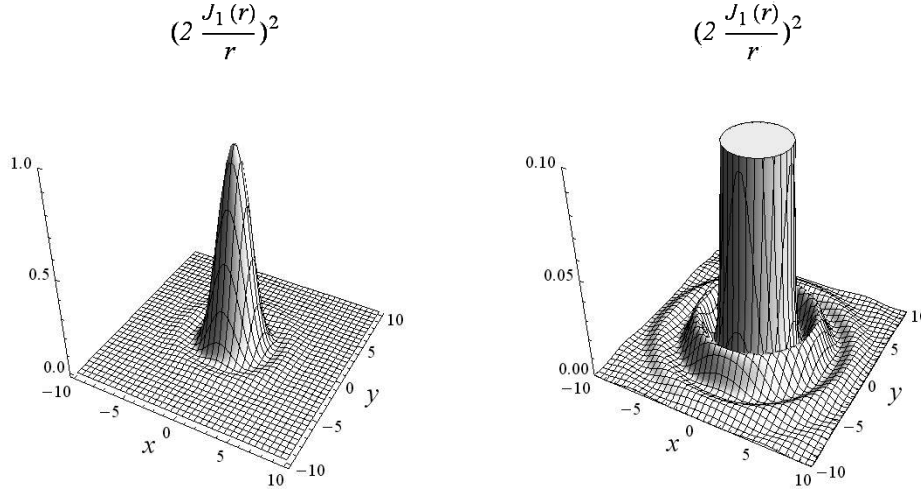


Figure 4.12 Point spread function from circular entrance pupil.

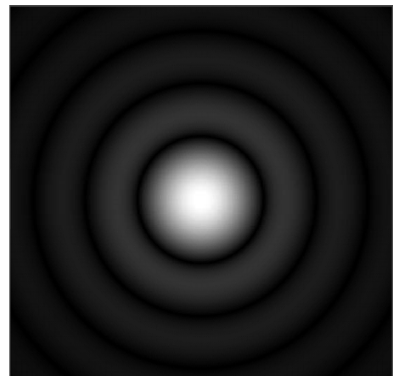


Figure 4.13 Simulated image of psf from circular entrance pupil.

Using (4.36) and (4.37), the psf for a circular aperture is

$$psf(r) = \left(\frac{2J_1\left(\frac{\pi D r}{\lambda f}\right)}{\frac{\pi D r}{\lambda f}} \right)^2 \tag{4.38}$$

A graph of $2\,J_1(x)$ is shown in Figure 4.11. From this figure we see that the first zero of $J_1(x)$ is at 3.83. Let r_1 denote the radius of the bright central spot corresponding to the first zero. Then

$$\frac{\pi D r_1}{\lambda f} = 3.83 \qquad \Rightarrow \quad r_1 = \frac{3.83}{\pi}\lambda\frac{f}{D} = 1.22\ \lambda F/\# \tag{4.39}$$

A graph of the psf from a circular entrance pupil is shown in Figure 4.12. The right side of the figure is shown with a magnified scale to illustrate the weak rings barely visible in the left side of the figure.

A simulated image of the psf from a circular entrance pupil is shown in Figure 4.13. The bright high-irradiance central area in this figure bounded by the black circle is called the *Airy disk* after Sir George Airy (1801–1892) who first derived (4.38). From (4.39), the diameter of the Airy disk is 2.4 λ F /#.

Although the results in this section were derived with the assumption that a simple thin lens formed the Fraunhofer diffraction pattern, the results are often an adequate approximation for estimating the diffraction spot size for fielded thick lens optics comprised of multiple lenses.

4.6 Thin Lens Diffraction MTF

Modulation and Modulation Transfer Function

Suppose the optical system is looking at a sinusoidal pattern that varies in the x direction but which is constant in the y direction. Then for a *psf* with zero extent, the radiance in the (x,y) image plane of Figure 4.1 would be given by

$$L_{ideal}(x) = L_0 + L_1 cos(2\pi\xi x) \tag{4.40}$$

which is illustrated in Figure 4.14. Although we have written this equation in terms of radiance we could have equally well written it in terms of irradiance E. Let $L_{\max}$ and $L_{\min}$ denote the maximum and minimum radiance in the image plane. Then $L_{\max} = L_0 + L_1$ and $L_{\min} = L_0 - L_1$. The modulation M_{ideal} of the wave in the x direction is defined by

$$M_{ideal} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} = \frac{L_0 + L_1 - (L_0 - L_1)}{L_0 + L_1 + (L_0 - L_1)} = \frac{L_1}{L_0} \tag{4.41}$$

Here the subscript *ideal* is used because this is the modulation observed in the image plane by an optical system with the ideal delta function *psf*.

The point spread function with a circular pupil is a symmetric function of x so the *lsf* associated with the *psf* of a circular pupil is an even function of x. Without loss of generality, we choose the normalization of $lsf(x)$ so that the area under *lsf* is one. The actual radiance pattern L_{act} seen in the image plane is the convolution of ideal radiance pattern convolved with the *lsf*:

$$\begin{aligned} L_{act}(x) &= L_{ideal}(x) * lsf(x) = \int L_{ideal}(x - x')\, lsf(x')\, dx' \\ &= L_0 \int lsf(x')\, dx' + L_1 \int lsf(x')\, cos(2\pi\xi(x - x'))\, dx' \\ &= L_0 + L_1 \int lsf(x') \left\{ cos(2\pi\xi x)\, cos(2\pi\xi x') + sin(2\pi\xi x)\, sin(2\pi\xi x') \right\} dx' \\ &= L_0 + L_1\, cos(2\pi\xi x) \int lsf(x')\, cos(2\pi\xi x')\, dx' \\ &= L_0 + L_1 \mathcal{F}\{lsf(x)\}\, cos(2\pi\xi x) \end{aligned} \tag{4.42}$$

The second line is obtained from the first line using (4.40). In the third line we used the normalization of $lsf(x)$ and expanded *cos* using a trigonometric identity. In the fourth line, the even symmetry of *lsf* and the odd symmetry of *sin* were used to eliminate the integral that included $lsf(x')$ and $sin\ (2\ \pi\ \xi\ x')$. In the fifth line we recognized that for an even function the *cos* transform of *lsf* is equivalent to the 1D Fourier transform of the *lsf*.

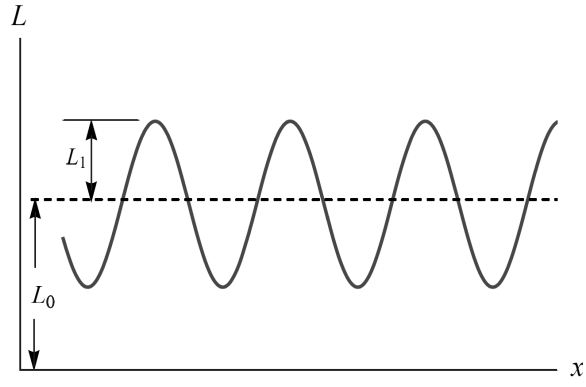


Figure 4.14 Defining the modulation transfer function MTF.

Using (4.41) and (4.42), the observed modulation M_{obs} with the assumed even *lsf* is

$$M_{obs} = \frac{L_1 \mathcal{F}\{lsf(x)\}}{L_0} \quad \text{assuming } lsf(-x) = lsf(x) \tag{4.43a}$$

If we relax the condition that *lsf* is an even function (this is appropriate for an arbitrary shaped entrance pupil), then we obtain a generalization of the above equation:

$$M_{obs} = \frac{L_1 \left|\mathcal{F}\{lsf(x)\}\right|}{L_0} \tag{4.43b}$$

The MTF is defined by

$$\text{MTF}(\xi) = \frac{M_{obs}}{M_{ideal}} = \frac{\dfrac{L_1 \left|\mathcal{F}\{lsf(x)\}\right|}{L_0}}{\dfrac{L_1}{L_0}}$$

$$= \left|\mathcal{F}\{lsf(x)\}\right| \text{ for wave varying only in the } x \text{ direction} \tag{4.44a}$$

Similarly,

$$\text{MTF}(\eta) = \left|\mathcal{F}\{lsf(y)\}\right| \text{ for wave varying only in the } y \text{ direction} \tag{4.44b}$$

In (4.44), the MTF was a function of ξ or η alone because the wave that was being modulated was varying in just one direction. The generalization of (4.44) to a wave that varies simultaneously with x and y is

$$\text{MTF}(\xi, \eta) \equiv \left|\mathcal{F}\mathcal{F}\{psf(x, y)\}\right| \tag{4.44c}$$

Incoherent Diffraction MTF

Using (4.36), the incoherent psf from a thin lens at the focal plane is given by

$$psf(x, y) = A\left(\left|\mathcal{F}\mathcal{F}\left\{p(x_a, y_a)\right\} /. \left\{\xi \to \frac{x}{\lambda f}, \eta \to \frac{y}{\lambda f}\right\}\right|\right)^2 \tag{4.45}$$

where A is a constant equal to the denominator of (4.36). We will later normalize the MTF so that $\text{MTF}(\xi,\eta) /. \{\xi \to 0, \eta \to 0\} = 1$ and for this reason set the normalization constant A to one.

Recall the following results from Table 2.3:

$$\mathcal{F}\mathcal{F}\{p(x_a, y_a)p^*(x_a, y_a)\} = P(\xi, \eta) \bigstar\bigstar P^*(\xi, \eta) \tag{4.46a}$$

$$\mathcal{F}\mathcal{F}\{p(x_a, y_a)p^*(x_a, y_a)\} = P(\xi, \eta) ** P^*(-\xi, -\eta) \tag{4.46b}$$

So

$$P(\xi, \eta) ** P^*(-\xi, -\eta) = P(\xi, \eta) \bigstar\bigstar P^*(\xi, \eta) \tag{4.46c}$$

Using Table 2.3 and the result $\mathcal{FF}\{G(\xi,\eta) = g(-x, -y)\}$, we obtain

$$\mathcal{FF}\left\{P\left(\frac{x}{a},\frac{y}{b}\right)\right\} = |a\ b|\,p(a\xi,-b\eta) \tag{4.47a}$$

$$\mathcal{FF}\left\{P^*\left(\frac{x}{a},\frac{y}{b}\right)\right\} = |a\ b|\,p^*(a\xi,b\eta) \tag{4.47b}$$

With A set equal to one, (4.45) becomes

$$psf(x,y) = P\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right) P^*\left\{\frac{x}{\lambda f},\frac{y}{\lambda f}\right\} \tag{4.48}$$

Using (4.44c)

$$\begin{aligned}
\mathrm{MTF}(\xi,\eta) &= \mathcal{FF}\{psf(x,y)\} \\
&= \mathcal{FF}\left\{P\left(\frac{x}{\lambda f},\frac{y}{\lambda f}\right) P^*\left\{\frac{x}{\lambda f},\frac{y}{\lambda f}\right\}\right\} \\
&= \lambda^4 f^4 p(-\lambda f\xi,-\lambda f\eta) ** p^*(\lambda f\xi,\lambda f\eta) \\
&= \lambda^4 f^4 p(\lambda f\xi,\lambda f\eta) \star\star p(\lambda f\xi,\lambda f\eta) \\
&\propto p(\lambda f\xi,\lambda f\eta) \star\star p(\lambda f\xi,\lambda f\eta)
\end{aligned} \tag{4.49}$$

The second line follows from (4.48); the third line follows from (4.47); the fourth line follows from (4.46c). Equation (4.49) is an important result and shows that for a thin lens the incoherent MTF is proportional to the autocorrelation function of the entrance pupil. The normalization constant is found from the requirement that at $\xi = \eta = 0$ the MTF is one.

From (4.35b), realize that the pupil function is either one or zero depending on whether or not (x_a,y_a) is inside or outside the pupil. Using (2.23) and (4.35b), note that the autocorrelation of a pupil function with itself is the area where the pupils overlap when they are displaced by an amount s. Let A_0 and A_{0L} denote the area of the pupil and the overlap area of the pupil when it is displaced by an amount s. Suppose the pupils are displaced in the x direction as shown in Figure 4.15. Then (4.49) implies

$$\mathrm{MTF}(\xi) = \frac{A_{0L}(s)}{A_0} /.\ s \to 10^{-3}\,\lambda f\xi \text{ where } \xi \text{ is in cycle/mm} \tag{4.50a}$$

$$\mathrm{MTF}(\xi) = \frac{A_{0L}(s)}{A_0} /.\ s \to \lambda\xi \text{ where } \xi \text{ is in cycle/mr} \tag{4.50b}$$

In (4.50), the wavelength is expressed in micrometers and the focal length is expressed in millimeters. If in (4.50), the displacement s was in the y direction, then we would replace ξ in these equations with η. A similar procedure is used to calculate the MTF in an arbitrary direction.

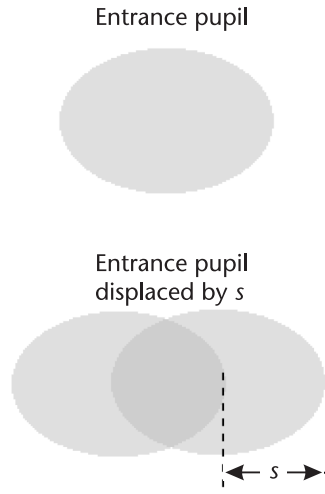


Figure 4.15 Method for calculating incoherent MTF.

Coherent Diffraction MTF

For a coherent system the MTF measures how the spatial frequencies of the optical disturbance *amplitude* are filtered by the optical system. In (4.35a), for x,y near the optical axis

$$\frac{k(x^2+y^2)}{2f} \ll \frac{2\pi}{\lambda f}(x\,x_a + y\,y_a)$$

that is, for small x and small y, the square of a term is negligible when compared to the term. Using this approximation, (4.35a) becomes

$$u(x,y) = Ae^{jkn\Delta_0}\frac{e^{jkf}}{f}P(\xi,\eta)/.\left\{\xi \to \frac{x}{\lambda f}, \eta \to \frac{y}{\lambda f}\right\}$$

$$\propto P\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right) \tag{4.51}$$

The coherent MTF is the Fourier transform of $u(x,y)$

$$\mathrm{MTF}_{coherent} \propto \mathcal{FF}\left\{P\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right)\right\}$$

$$\propto p(-\lambda f\xi, -\lambda f\eta) \tag{4.52a}$$

where we have used 2D extension of (2.47) to realize that the Fourier transform of the pupil function done twice yields the pupil function with negative arguments. At zero frequency the coherent MTF is one. Thus, providing there is no obstruction when $x_a = y_a = 0$, then

$$\mathrm{MTF}_{coherent} = p(-\lambda f\xi, -\lambda f\eta) \tag{4.52b}$$

Frequently the pupil function is symmetric through the origin so $p(-x_a,-y_a) = p(x_a,y_a)$. Hence, (4.52b) becomes

$$\mathrm{MTF}_{coherent} = p(\lambda f\xi, \lambda f\eta) \tag{4.52c}$$

4.7 Calculating Diffraction MTF with Pencil and Paper

Circular Pupil: Coherent MTF

For a circular pupil the pupil function is expressed by

$$p(r_a) = cyl\left(\frac{r_a}{D}\right)$$

where D is the diameter of the pupil and r_a is a radius variable in the aperture plane. Using (4.52),

$$\mathrm{MTF}_{coherent} = cyl\left(\frac{\lambda f\rho}{D}\right) = cyl\left(\frac{\rho}{2\rho_{oco}}\right) \tag{4.53a}$$

where ρ is the radial spatial frequency in cycles per unit distance, f is the focal length, λ is the radiation wavelength, and ρ_{oco} is the coherent optical cutoff frequency as follows:

$$\rho_{oco} = \frac{1}{2}\frac{D}{\lambda f} \tag{4.53b}$$

Circular Pupil: Incoherent MTF

It is an exercise in geometry to show that the overlap area in Figure 4.16 is given by

$$A_{OL}(s) = 2\left(\frac{D}{2}\right)^2\left[\cos^{-1}\left(\frac{s}{D}\right) - \frac{s}{D}\sqrt{1-\left(\frac{s}{D}\right)^2}\right] \quad \text{for } s \le D$$

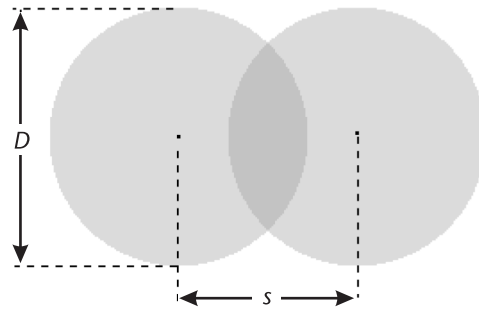


Figure 4.16 Incoherent MTF calculation for circular entrance pupil.

When $s = 0$, the overlap area is the area of a circle

$$A_0 = \pi\left(\frac{D}{2}\right)^2$$

Using (4.50),

$$\begin{aligned}\text{MTF}(\xi) &= \frac{2}{\pi}\left(cos^{-1}\left(\frac{s}{D}\right) - \frac{s}{D}\sqrt{1-\left(\frac{s}{D}\right)^2}\right) /. \, s \to \lambda f \xi \\ &= \frac{2}{\pi}\left(cos^{-1}\left(\frac{\lambda f \xi}{D}\right) - \frac{\lambda f \xi}{D}\sqrt{1-\left(\frac{\lambda f \xi}{D}\right)^2}\right) \\ &= \frac{2}{\pi}\left(cos^{-1}\left(\frac{\xi}{\xi_{oco}}\right) - \frac{\xi}{\xi_{oco}}\sqrt{1-\left(\frac{\xi}{\xi_{oco}}\right)^2}\right) \end{aligned} \tag{4.54a}$$

where ξ_{oco} is the incoherent optical cutoff frequency:

$$\xi_{oco} = \frac{D}{\lambda f} \tag{4.54b}$$

Here we have done the calculation for displacements in the x direction but because of the symmetry of the circular aperture the same result is obtained for displacements in either the y direction or radial direction. Comparing (4.54b) with (4.53b), we note that the incoherent optical cutoff frequency is twice the coherent optical frequency.

4.8 Programs for Calculating Incoherent Diffraction MTF

CODE V has the ability to numerically calculate the incoherent diffraction MTF. In versions of Mathematica 6.0 and higher it is easy to write programs that calculate incoherent diffraction MTF analytically as well as numerically. Mathematica is a general-purpose computer language with capabilities that overlap those of MATLAB. For any physically realizable pupil shape, one can numerically compute the diffraction MTF but for some aperture shapes Mathematica may not succeed in the analytical calculation. An ability to numerically calculate diffraction MTF is useful for checking the correctness of analytical calculations and is a necessity when the analytical calculation fails. For these reasons it is useful to have programs that calculate diffraction MTF numerically and analytically.

One common way for expressing MTF is in terms of a cutoff frequency as was done in (4.54). In Figure 4.15, for all s values greater than $s_{\max}$, the overlap area A_{0L} is zero. The $s_{\max}$ values in x and y directions are often different and this implies the cutoff frequencies in these directions are different. Optical cutoff frequencies

f_{OCO} in the x and y directions are defined in terms of $s_{\max}$ and are commonly expressed in cycles per milliradian or cycles per millimeter:

$$f_{oco} = \frac{s_{\max}}{\lambda} \qquad \text{[cycle/mr]} \tag{4.55a}$$

$$f_{oco} = \frac{s_{\max}}{10^{-3}\lambda f} \qquad \text{[cycle/mm]} \tag{4.55b}$$

In (4.55), displacement s_{max} and focal length f are expressed in millimeters and the wavelength λ is expressed in micrometers. The dimensionless normalized frequency f_n is defined by

$$f_n = \frac{f_s}{f_{oco}} \tag{4.56}$$

where f_s is a spatial frequency.

Besides expressing MTF in terms of a normalized frequency, there are two other common ways to express MTF: in one method the independent variable is cycles per millimeter; in the other method the independent variable is cycles per milliradian. MTF expressed in cycles per milliradian is most useful for analyzing the field effectiveness of a sensor. Lens designers, however, often prefer to have MTF expressed in cycles per millimeter. Although a MTF can be computed in any direction, frequently the MTF is represented by a MTF in either the horizontal or vertical directions.

As indicated in Table 4.2, 12 programs are provided to enable the reader to calculate MTF numerically or analytically in the horizontal or vertical direction in units of normalized frequency, cycles per millimeter, or cycles per milliradian. The programs have been validated [1] by internal consistency checks and by comparison with Code V calculations.

Table 4.2 Programs for computing incoherent diffraction MTF.

No	Program Name	Anal/Num	Frequency	Direction
1	MTFNormalizedEqHor	Analytical	Normalized	Horizontal
2	MTFNormalizedEqVer	Analytical	Normalized	Vertical
3	NMTFNormalizedHor	Numerical	Normalized	Horizontal
4	NMTFNormalizedVer	Numerical	Normalized	Vertical
5	MTFCyclesPerMmEqHor	Analytical	Cycles/ mm	Horizontal
6	MTFCyclesPerMmEqVer	Analytical	Cycles/ mm	Vertical
7	NMTFCyclesPerMmHor	Numerical	Cycles/ mm	Horizontal
8	NMTFCyclesPerMmVer	Numerical	Cycles/ mm	Vertical
9	MTFCyclesPerMrEqHor	Analytical	Cycles/ mr	Horizontal
10	MTFCyclesPerMrEqVer	Analytical	Cycles/ mr	Vertical
11	NMTFCyclesPerMrHor	Numerical	Cycles/ mr	Horizontal
12	NMTFCyclesPerMrVer	Numerical	Cycles/ mr	Vertical

Programs 1–4 (Normalized Frequency)

Program 1. MTFNormalizedEqHor takes as input the normalized aperture shape and outputs equations that describe the normalized MTF in the horizontal direction. The normalized aperture shape is the shape of the aperture with the maximum dimension in the horizontal direction scaled to one.

```
MTFNormalizedEqHor[Ineq_, {x_, y_}, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞},
     Assumptions → s ≥ 0];
  temp /. s → f]
```

Program 2. MTFNormalizedEqVer takes as input the normalized aperture shape and outputs equations that describe the normalized MTF in the vertical direction. The normalized aperture shape is the shape of the aperture with the maximum direction in the vertical direction normalized to one.

```
MTFNormalizedEqVer[Ineq_, {x_, y_}, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. y → (y - s))], {y, -∞, ∞}, {x, -∞, ∞}, Assumptions → s ≥ 0];
  temp /. s → f]
```

Program 3. NMTFNormalizedHor takes as input the normalized aperture shape and outputs a table that describes the normalized MTF in the horizontal direction.

```
NMTFNormalizedHor[Ineq_, {x_, y_}, NPoints_] := Module[{A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞}]},
    {s, 0, 1, 1/NPoints}] // Chop
]
```

Program 4. NMTFNormalizedVer takes as input the normalized aperture shape and outputs a table that describes the normalized MTF in the vertical direction.

```
NMTFNormalizedVer[Ineq_, {x_, y_}, NPoints_] := Module[{A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. y → (y - s))], {x, -∞, ∞}, {y, -∞, ∞}]},
    {s, 0, 1, 1/NPoints}] // Chop
]
```

Programs 5–8 (Cycles per Millimeter)

Program 5. MTFCyclesPerMmEqHor takes as input the aperture shape, given as an inequality in units of millimeters, and outputs equations that describe the MTF in the horizontal direction in units of cycles per millimeter.

```
MTFCyclesPerMmEqHor[Ineq_, {x_, y_}, λ_, fl_, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞},
     Assumptions → s ≥ 0];
  (* s is in mm, λ is in μ, fl is focal length in mm, f is in cycles/mm
   Factor of 10^-3 changes μ to mm *)
  temp /. s → 10^-3 λ fl f]
```

Program 6. MTFCyclesPerMmEqVer takes as input the aperture shape, given as an inequality in units of millimeters, and outputs equations that describe the MTF in the vertical direction in units of cycles per millimeter.

```
MTFCyclesPerMmEqVer[Ineq_, {x_, y_}, λ_, fl_, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. y → (y - s))], {y, -∞, ∞}, {x, -∞, ∞},
     Assumptions → s ≥ 0];
  (* s is in mm, λ is in μ, fl is focal length in mm, f is in cycles/mm
   Factor of 10^-3 changes μ to mm *)
  temp /. s → 10^-3 λ fl f]
```

Program 7. NMTFCyclesPerMmHor does the same thing as Program 5, MTFCylesPerMmEqHor, only the calculation is done numerically so results are output as a table rather than as equations.

```
NMTFCyclesPerMmHor[Ineq_, {x_, y_}, λ_, fl_, smax_, NPoints_] := Module[{temp, A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞}]},
     {s, 0, smax, smax/NPoints}] // Chop;
  (* s is in mm, λ is in μ, fl is focal length in mm, f is in cycles/mm
   Factor of 10^-3 changes μ to mm *)
  temp /. {a_, b_} → {a/(10^-3 λ fl), b}
]
```

Program 8. NMTFCyclesPerMmVer does the same thing as Program 6, MTFCyclesPerMmEqVer, only the calculation is done numerically so results are output as a table rather than as equations.

```
NMTFCyclesPerMmVer[Ineq_, {x_, y_}, λ_, fl_, smax_, NPoints_] := Module[{temp, A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. y → (y - s))], {y, -∞, ∞}, {x, -∞, ∞}]},
     {s, 0, smax, smax/NPoints}] // Chop;
  (* s is in mm, λ is in μ, fl is focal length in mm, f is in cycles/mm
   Factor of 10^-3 changes μ to mm *)
  temp /. {a_, b_} → {a/(10^-3 λ fl), b}
]
```

Programs 9–12 (Cycles per Milliradian)

Program 9. MTFCyclesPerMrEqHor takes as input the aperture shape and outputs equations that describe the MTF in the horizontal direction in units of cycles per milliradian.

```
MTFCyclesPerMrEqHor[Ineq_, {x_, y_}, λ_, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞},
     Assumptions → s ≥ 0];
  (* s is in mm, λ is in μ, f is in cycles/mr *)
  temp /. s → λ f]
```

Program 10. MTFCyclesPerMrEqVer takes as input the aperture shape and outputs the equations that describe the MTF in the vertical direction in units of cycles per milliradian.

```
MTFCyclesPerMrEqVer[Ineq_, {x_, y_}, λ_, f_] := Module[{temp, A0, s},
  A0 = Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = 1/A0 Integrate[Boole[Ineq && (Ineq /. y → (y - s))], {y, -∞, ∞}, {x, -∞, ∞},
     Assumptions → s ≥ 0];
  (* s is in mm, λ is in μ,  f is in cycles/mm *)
  temp /. s → λ f]
```

Program 11. NMTFCyclesPerMrHor does the same calculation as Program 9, MTFCyclesPerMrEqHor, only the calculation is done numerically so the results are output as a table instead of as equations.

```
NMTFCyclesPerMrHor[Ineq_, {x_, y_}, λ_, smax_, NPoints_] := Module[{temp, A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞}]},
     {s, 0, smax, smax/NPoints}] // Chop;
  (* s is in mm, λ is in μ, f is in cycles/mr *)
  temp /. {a_, b_} → {a/λ, b}]
```

Program 12. NMTFCyclesPerMrVer does the same calculation as Program 10, MTFCyclesPerMrEqVer, only the calculation is done numerically so the results are output as a table instead of as equations.

```
NMTFCyclesPerMrVer[Ineq_, {x_, y_}, λ_, smax_, NPoints_] := Module[{temp, A0, s},
  A0 = NIntegrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
  temp = Table[{s, 1/A0 NIntegrate[Boole[Ineq && (Ineq /. y → (y - s))], {y, -∞, ∞}, {x, -∞, ∞}]},
    {s, 0, smax, smax/NPoints}] // Chop;
  (* s is in mm, λ is in μ,  f is in cycles/mm *)
  temp /. {a_, b_} → {a/λ, b}]
```

By focusing on Program 9, **MTFCyclesPerMrEqHor** we provide a framework for understanding how all of the programs work. The pupil shape is input as an inequality called **Ineq** in the program. The function **Boole** has **Ineq** as an argument and returns 1 if the inequality is satisfied and 0 otherwise. Then

```
Integrate[Boole[Ineq], {x, -∞, ∞}, {y, -∞, ∞}];
```

in the second line of the program integrates over the entire plane and returns an analytical expression for the area of the pupil. Similarly, the second line of the program,

```
Integrate[Boole[Ineq && (Ineq /. x → (x - s))], {x, -∞, ∞}, {y, -∞, ∞}, Assumptions → s ≥ 0];
```

does a calculation of the overlap area of the pupil with the pupil displaced in the *x-direction*; the last line of the code calculates the MTF using (4.50b).

Note that Program 11, **NMTFCyclesPerMrHor** does the same calculation as Program 9 only it uses the *Mathematica* function **NIntegrate** instead of **Integrate**. **NIntegrate** does the calculation numerically, so in the second line of Program 11 we need to create a table of values and this is done with the *Mathematica* **Table** function.

We now illustrate the use of the programs with some examples.

Example 4.3

Calculate MTF expressed in normalized frequency for the circular, square, and semicircular entrance pupils illustrated in Figure 4.17.

For the circular pupil, the input

```
MTFNormalizedEqHor[x^2 + y^2 < (1/2)^2, {x, y}, ξn] // Simplify // TraditionalForm
```

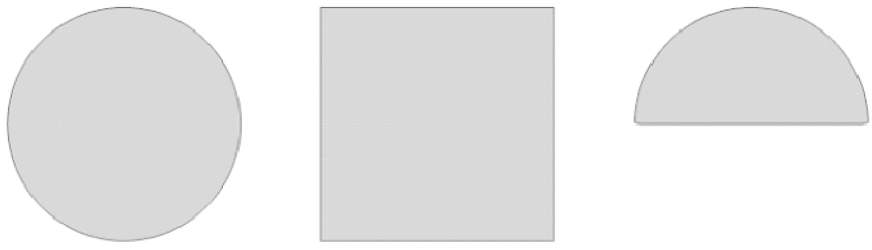


Figure 4.17 Shapes used to illustrate the use of the **MTFNormalizedEqHor** program.

yields the analytical result for the incoherent MTF:

$$\begin{cases} 1 & \xi_n = 0 \\ \dfrac{2\left(\cos^{-1}(\xi_n) - \xi_n\sqrt{1-\xi_n^2}\right)}{\pi} & 0 < \xi_n < 1 \end{cases}$$

The first argument of **MTFNormalizedEqHor** is an inequality that defines a circular pupil. The radius was chosen to be 1/2 because **MTFNormalizedEqHor** requires the maximum dimension in the horizontal direction to be one. The second argument tells *Mathematica* the variables in the pupil function that are to be integrated over. The symbol for the third argument is chosen at the convenience of the analyst and is here chosen to be ξ_n to indicate normalized spatial frequencies in the horizontal direction. The commands **//Simplify//TraditionalForm** at the end of the input tells *Mathematica* to simplify the result and express the output in the traditional form favored by the general science and engineering community.

For the square pupil the input

```
MTFNormalizedEqHor[Abs[x] < 1/2 && Abs[y] < 1/2, {x, y}, ξn]
```

yields the analytical result

$$\begin{cases} 1-\xi_n & 0 \le \xi_n < 1 \\ 0 & \text{True} \end{cases}$$

For the semicircular pupil, the input

```
MTFNormalizedEqHor[x^2 + y^2 < (1/2)^2 && y > 0, {x, y}, ξn] // Simplify // TraditionalForm
```

yields the analytical MTF result for frequencies in the x direction

$$\begin{cases} 1 & \xi_n = 0 \\ \dfrac{2\left(\cos{-1}(\xi_n) - \xi_n\sqrt{1-\xi_n^2}\right)}{\pi} & 0 < \xi_n < 1 \end{cases}$$

which we note is the same expression as the circular pupil in the x direction.

It is instructive to compare the analytical results shown above with numerical calculations. The commands

```
NMTFNormalizedHor[x^2 + y^2 < (1/2)^2, {x, y}, 30.]

NMTFNormalizedHor[Abs[x] < 1/2 && Abs[y] < 1/2, {x, y}, 30.]

NMTFNormalizedHor[x^2 + y^2 < (1/2)^2 && y > 0, {x, y}, 30.]
```

generate tables, each with 30 points, that describe the MTF with normalized frequencies for a circular, square, and semicircular pupil. Graphs comparing the analytical and numerical calculations are shown in Figure 4.18.

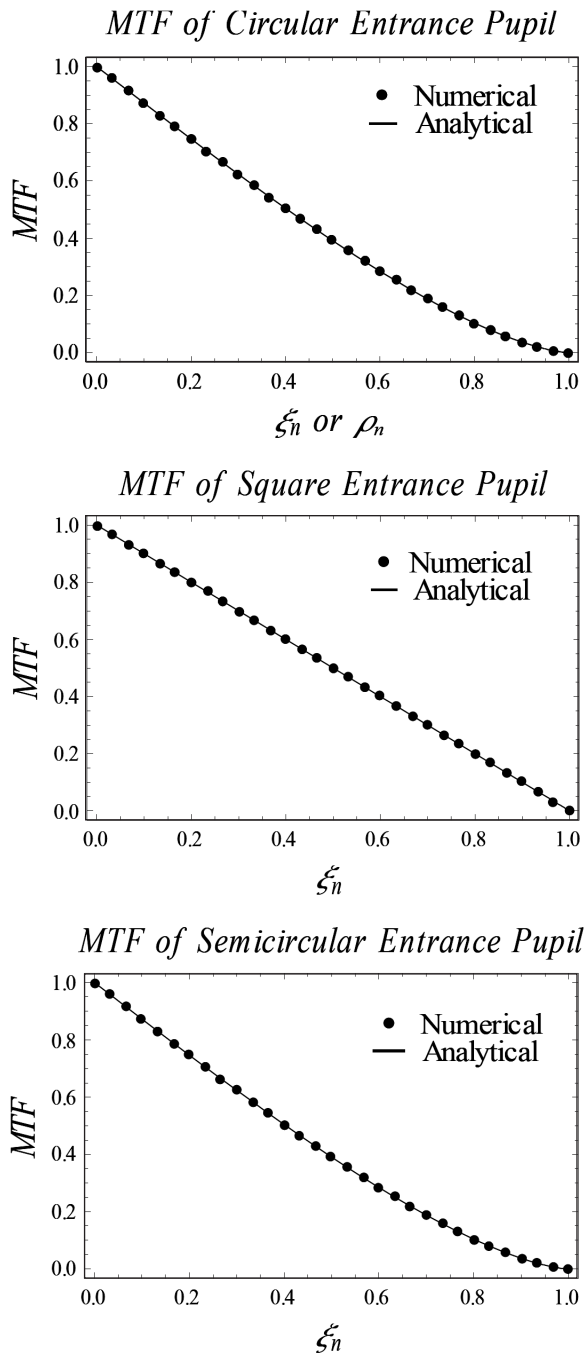


Figure 4.18 Comparison of analytical and numerical calculations for circular, square, and semicircular entrance pupils.

The preceding example showed how to calculate diffraction MTF in terms of normalized spatial frequency even without specific dimensions. The next example shows how to calculate diffraction MTF in terms of cycles per millimeter for an entrance aperture with specific dimensions.

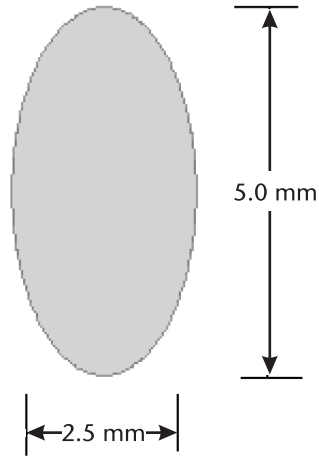


Figure 4.19 Elliptical entrance pupil.

Example 4.4

For the illustrated elliptical entrance pupil, calculate the incoherent MTF assuming radiation with a central wavelength of 10 μm assuming the lens has a focal length of 10 mm.

The following commands output equations and numerical tables for the horizontal and vertical MTF for the elliptical aperture shown in Figure 4.19.

$$\text{MTFCyclesPerMmEqHor}\left[\left(\frac{x}{1.25}\right)^2+\left(\frac{y}{2.5}\right)^2<1,\{x,y\},10,10,\xi\right]$$

$$\text{NMTFCyclesPerMmHor}\left[\left(\frac{x}{1.25}\right)^2+\left(\frac{y}{2.5}\right)^2<1,\{x,y\},10,10,2.5,30.\right]$$

$$\text{MTFCyclesPerMmEqVer}\left[\left(\frac{x}{1.25}\right)^2+\left(\frac{y}{2.5}\right)^2<1,\{x,y\},10,10,\eta\right]$$

$$\text{NMTFCyclesPerMmVer}\left[\left(\frac{x}{1.25}\right)^2+\left(\frac{y}{2.5}\right)^2<1,\{x,y\},10,10,5.0,30.\right]$$

The first argument always specifies the pupil shape. In the second and fourth line above the last argument is 30, which specifies the MTF is to be calculated at 30 frequencies uniformly distributed between the zero and the cutoff frequency. The next to the last arguments of 2.5 and 5.0 correspond to the maximum s values in the horizontal and vertical directions; greater displacements correspond to zero overlap area. Space limitations preclude exhibiting the equations and tables produced by these commands. Graphs comparing the analytical and numerical calculations are shown in Figure 4.20.

Note that for the elliptical pupil of Figure 4.19, the cutoff frequency is a factor of 2 higher in the vertical direction than in the horizontal direction and this corresponds to the elliptical aperture being twice as long as it is wide.

The next example demonstrates how to calculate the diffraction MTF for a distributed aperture.

Example 4.5

For the distributed entrance pupil shown in Figure 4.21, calculate the incoherent MTF in the vertical and horizontal directions assuming a focal length of 10 mm and a central wavelength of 10 μm.

By symmetry, the MTFs in the vertical and horizontal directions are the same. The commands used to define the distributed aperture are as follows:

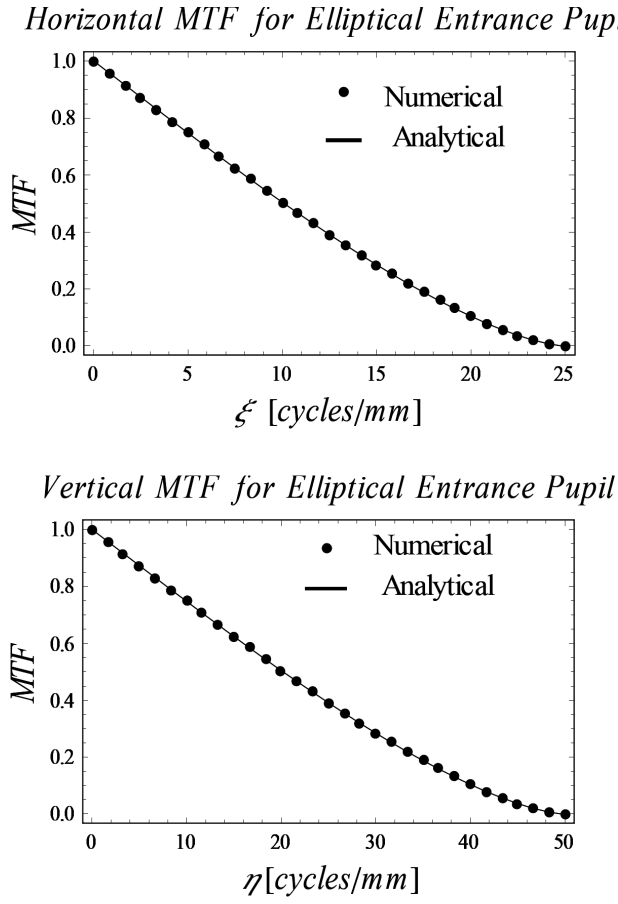


Figure 4.20 MTF for the elliptical entrance pupil.

```
Region1 = x^2 + (y - 1.9)^2 < 0.6^2;
Region2 = x^2 + (y + 1.9)^2 < 0.6^2;
Region3 = (x - 1.9)^2 + y^2 < 0.6^2;
Region4 = (x + 1.9)^2 + y^2 < 0.6^2;
Region = Region1 || Region2 || Region3 || Region4;
```

The four inequalities define each of the four subapertures. The last line defines the distributed aperture by repeated use of the OR command, represented in *Mathematica* by ||. We can check that the distributed aperture defined in the last line is what was wanted by using the RegionPlot command,

```
RegionPlot[Region, {x, -3, 3}, {y, -3, 3}, AspectRatio → Automatic, Frame → False]
```

which was used to generate Figure 4.21. The following commands used to generate the equation and table that define the horizontal or vertical MTF:

```
MTFCyclesPerMmEqVer[Region, {x, y}, 10., 10., η]
```

```
NMTFCyclesPerMmVer[Region, {x, y}, 10., 10., 5, 30.]
```

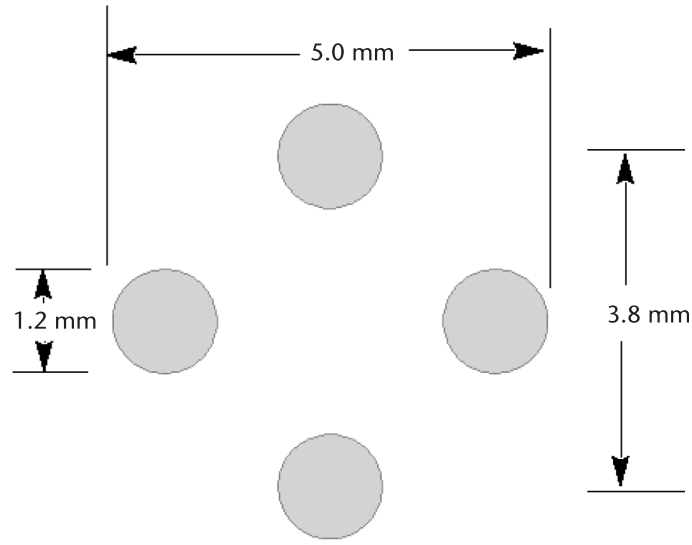


Figure 4.21 Distributed aperture.

Space limitations preclude exhibiting the equation or table produced by the above commands. A comparison of the analytical and numerical results generated by the above commands is shown in Figure 4.22.

Note that this particular distributed aperture does not pass frequencies between approximately 12 and 26 cycle/mm. The reason for this is that as the pupil is displaced there are regions where there is no overlap and these correspond to spatial frequencies where the MTF is zero.

The next example is included because an entrance pupil with this shape is being considered for use in a military sensor.

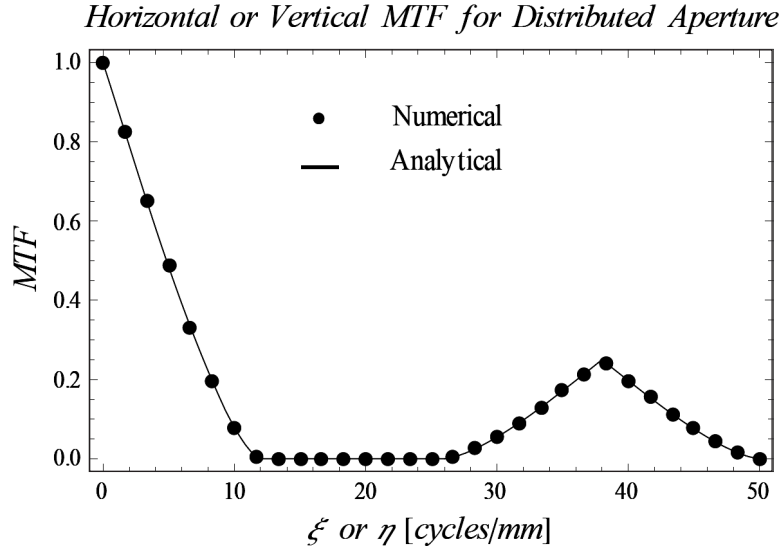


Figure 4.22 Horizontal or vertical MTF for distributed aperture.

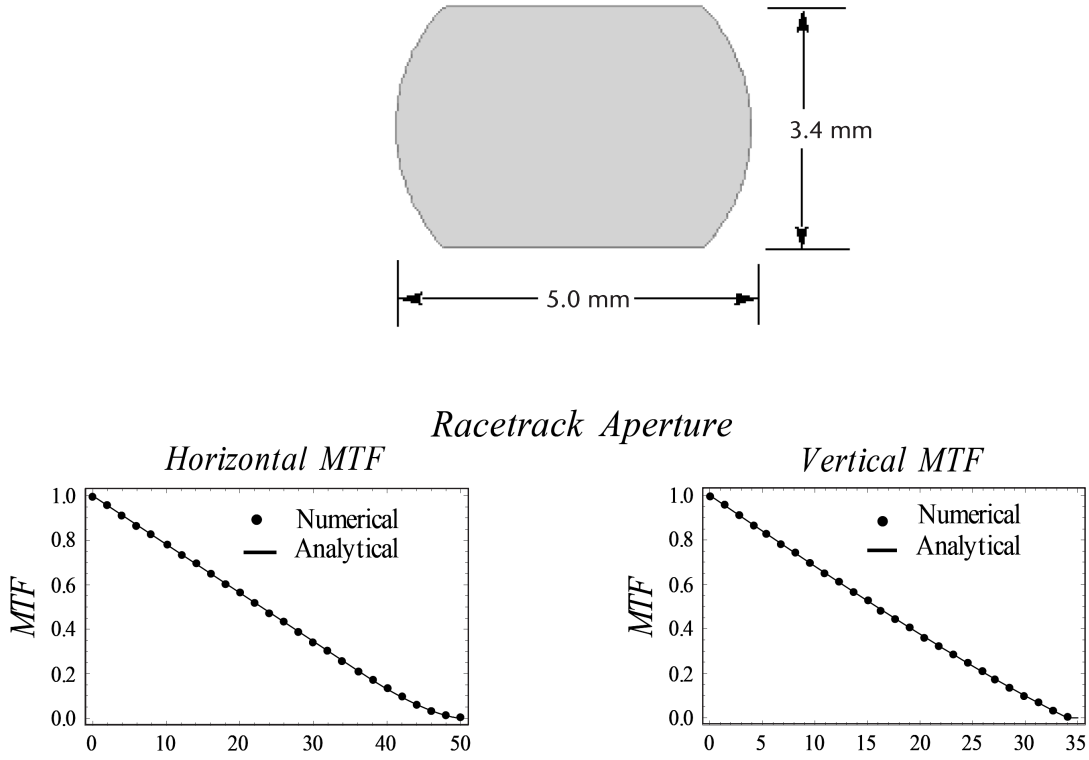


Figure 4.23 Racetrack entrance pupil with horizontal and vertical MTFs.

Example 4.6

For the entrance pupil shown in Figure 4.23, calculate the incoherent MTF in the vertical and horizontal directions assuming a focal length of 10 mm and a central wavelength of 10 μm.

The following commands output equations and numerical tables for the horizontal and vertical MTFs:

```
MTFCyclesPerMmEqHor[x^2 + y^2 < 2.5^2 && Abs[y] < 1.7, {x, y}, 10, 10, ξ]

MTFCyclesPerMmEqVer[x^2 + y^2 < 2.5^2 && Abs[y] < 1.7, {x, y}, 10, 10, η]

NMTFCyclesPerMmHor[x^2 + y^2 < 2.5^2 && Abs[y] < 1.7, {x, y}, 10., 10., 5.0, 25.]

NMTFCyclesPerMmVer[x^2 + y^2 < 2.5^2 && Abs[y] < 1.7, {x, y}, 10., 10., 3.4, 25.]
```

Space limitations preclude exhibiting the equations or tables generated by these commands. Figure 4.23 compares the analytical and numerical calculations.

4.9 Applications of Diffraction Theory

Frequency analysis of optical systems

The coherent and incoherent MTF describes the filtering action of a lens on a scene. Material for calculating incoherent MTF are given in Sections 4.6, 4.7, and 4.8 in

this chapter. Additional or complementary material on the frequency analysis of optical systems is given in [5–9, 11, 12].

Application to geometric optics

As will be explained in Chapter 6, for a proposed lens design, a lens designer traces a variety of rays from a point on the object so that each ray passes through the aperture stop and onto the focal plane. The rays make a spot diagram on the focal plane that defines the geometrical optics psf. For imaging purposes, a smaller psf is better. One use of diffraction theory is that it tells the lens designer that when the geometrical psf is smaller than the Airy disk, little benefit is gained by increasing the complexity of the optical design. Examples of spot diagrams are given in [18].

Point spread function of distributed aperture

In Section 4.5 we showed how to calculate the diffraction psf from a lens for square and circular entrance pupils, but did not show how to do that for distributed apertures such as that shown in Figure 4.21. Equation (4.36) applies to distributed as well as simple apertures. To use this equation on the distributed aperture of Figure 4.21, represent the pupil function as the convolution of a circle (representing the subapertures) with four delta functions located at the center of each subaperture. Then the Fourier transform yields the product of the Fourier transform of the circle multiplied by the Fourier transform of the four delta functions. This is essentially the *array theorem*, which is useful for calculating the psf of a distributed aperture. The method of computing Fraunhofer diffraction patterns from an array of apertures is described in [6, 7, 9, 11, 17].

Optical image processing

In Section 4.4 we showed that when a plane monochromatic wave normally illuminates a transparency in the plane of a convex lens, then a focal length away from the lens we observe a Fraunhofer intensity pattern that consists of the Fourier transform of the transparency convolved with the Fourier transform of the pupil function. It is also true that if a transparency $\tau(x,y)$ located a distance f_t in front of a convex lens (the t subscript is short for *transforming*) is illuminated by a normally incident monochromatic plane wave, then the Fourier transform of $\tau(x,y)$ is produced a distance f_t behind the lens. This is termed the *Fourier transform plane*. By placing an opaque or partially opaque mask in the Fourier transform plane, we can eliminate or reduce selected frequencies associated with the Fourier transform of $\tau(x,y)$. Another lens placed a distance f_i behind the Fourier transform plane (the i subscript is for *imaging*) then produces a Fourier transform of the Fourier transform plane a distance f_i behind the second lens. Note that the Fourier transform of $\tau(x,y)$ done twice produces $\tau(-x, -y)$ a copy of the transparency inverted through the origin. So using this system we can produce the original transparency inverted through the origin but with selected spatial frequencies either eliminated or reduced. This is an example of an $f\,f - f\,f$ image processing system. Discussions of optical image processing are given in [7, 9, 11, 13, 17].

Stellar interferometry

In Section 4.2 we showed that (4.21) describes the spatial coherence length in terms of the wavelength and the angle subtended by the source. More precise equations similar to (4.21) (see Van Cittert-Zernike theorem) are used to estimate the diameter of stars. Discussions of how to measure the diameter of a star using spatial coherence concepts are given in [6, 7].

Apodization

The point spread function of circular entrance pupil was described in Section 4.5. From Figure 4.11 we observe that the psf has a central maximum and a series of weak rings. The diffraction rings make it difficult to see a star with weak luminance that has a small angular separation from a star with strong luminance. This problem can be addressed (1) by changing the shape of the aperture, (2) by having the transmission of the lens decline in the radial direction approaching the edge of the entrance pupil, and (3) by introducing a phase plate adjacent to the entrance pupil. The idea is to modify the diffraction point spread function with the view toward either eliminating the rings in Figure 4.12 or reducing them. Reducing the amplitude of the rings or eliminating the rings altogether generally comes at the expense of reduced irradiance in the image and by making the central peak broader. Apodization is described in [6, 7, 11].

Detector MTF from Fraunhofer diffraction pattern

Let $e(x,y)$, $d(x,y)$, and $s(x,y)$ denote the irradiance on the focal plane from a scene, the detector shape, and the signal produced by the detector, respectively. The function $d(x,y)$ has the property

$$d(x,y) = \begin{cases} 1 & \text{for } x,y \text{ in the detector} \\ 0 & \text{for } x,y \text{ outside the detector} \end{cases} \tag{4.57}$$

The signal is the convolution of $e(x,y)$ with $d(x,y)$

$$s(x,y) = e(x,y) ** d(x,y) \quad \text{so} \tag{4.58a}$$

$$S(\xi,\eta) = E(\xi,\eta)D(\xi,\eta)$$

where

$$E(\xi,\eta) \equiv \mathcal{FF}\{e(x,y)\},\ D(\xi,\eta) \equiv \mathcal{FF}\{d(x,y)\} \text{ and } S(\xi,\eta) \equiv \mathcal{FF}\{s(x,y)\} \tag{4.59}$$

The modulation transfer function of the detector is defined by

$$MTF_{det} \equiv \left|D(\xi,\eta)\right| \tag{4.60}$$

We will show how to obtain MTF_{det} from a Fraunhofer diffraction pattern produced by an aperture with the same shape as the detector. Chose an entrance pupil with the same shape and size as the detector $d(x, y)$, illuminate it with monochromatic radiation, and examine the diffraction psf or equivalently the Fraunhofer

diffraction pattern associated with the entrance pupil. From (4.36), we see that in this case

$$psf(x,y) \propto \left(\left| \mathcal{FF}\{d(x,y)\} /. \left\{ \xi \to \frac{x}{\lambda f}, \eta \to \frac{y}{\lambda f} \right\} \right| \right)^2$$

$$\propto |D(\xi,\eta)|^2 /. \left\{ \xi \to \frac{x}{\lambda f}, \eta \to \frac{y}{\lambda f} \right\} \tag{4.61}$$

From a comparison of (4.60) and (4.61), it is apparent that the detector MTF can be obtained from the psf using this procedure:

- Measure the Fraunhofer irradiance pattern produced by an aperture whose pupil function $p(x,y)$ has the same shape and size as $d(x,y)$.
- Normalize the Fraunhofer irradiance pattern so that it is one at the origin. This is $psf(x,y)$.
- Take the square root of $psf(x,y)$.
- Using (4.61) and the observation that $psf\ (0,0) = |\ D(0,0)| = 1$, we conclude that $\sqrt{psf(x,y)} /. \left\{ \frac{x}{\lambda f} \to \xi, \frac{y}{\lambda f} \to \eta \right\} = |\ D(\xi,\eta)\ |$.

By observing how the Fraunhofer diffraction pattern varies with x and y we can determine the detector MTF in any direction from the Fraunhofer diffraction pattern [2].

Example 4.7

Use Fraunhofer diffraction theory to evaluate the detector cutoff frequency in the x and y directions for the square and diamond-shaped detectors shown in Figure 4.24. Calculate and graph the detector MTFs in these directions.

For the square detector:

$$\mathrm{MTF}_{square}(\xi) = \frac{1}{a} \left| \mathcal{F} \left\{ rect\left(\frac{x}{a}\right) rect\left(\frac{y}{a}\right) \right\} \right|$$

$$= \left| \mathrm{s}inc(a\xi) \right| rect\left(\frac{y}{a}\right) = \left| \frac{sin(\pi a \xi)}{\pi a \xi} \right| rect\left(\frac{y}{a}\right) \tag{4.62a}$$

The cutoff frequency of the square detector is the lowest frequency for which the MTF is zero:

$$\xi_{square\,cutoff} = \frac{1}{a} \tag{4.62b}$$

Photographs of the Fraunhofer diffraction pattern from a square aperture are displayed in Hecht and Zajac [7] and are shown using computer graphics in Figure 4.25. The straight lines in the contour plot correspond to frequencies where the MTF is zero. From this figure we see that the cutoff frequency for the diamond-shaped detector will be greater by a factor $\sqrt{2}$ than that of the square detector

$$\xi_{diamond\,cutoff} = \frac{\sqrt{2}}{a} \tag{4.63a}$$

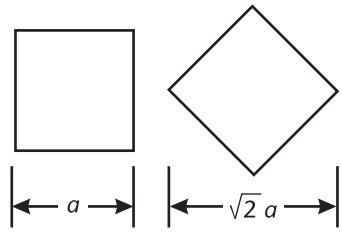
This encourages a detailed calculation of the detector MTF for a diamond-shaped detector which yields (M. Friedman, *Modulation Transfer Functions and Detector Shape*, unpublished report, June 1988):

$$\mathrm{MTF}_{diamond}(\xi) = \frac{sin^2\left(\frac{\pi a\ \xi}{\sqrt{2}}\right)}{\left(\frac{\pi a\ \xi}{\sqrt{2}}\right)^2} \tag{4.63b}$$

From the symmetry of the square detector, the MTFs in the vertical and horizontal directions are the same. Similarly, for the diamond-shaped detector the MTFs in the vertical and horizontal directions are the same. A comparison of the square and diamond MTFs is shown in Figure 4.24(b).

From Figure 4.24(b) it is clear that the diamond-shaped detector has a better MTF than the square detector in the horizontal and vertical directions. Observe from Figures 4.24(b) and 4.25 that there is a phase reversal above the cutoff frequency for the square detector that is not present for the diamond-shaped detector. A performance model for

(a)



(b)

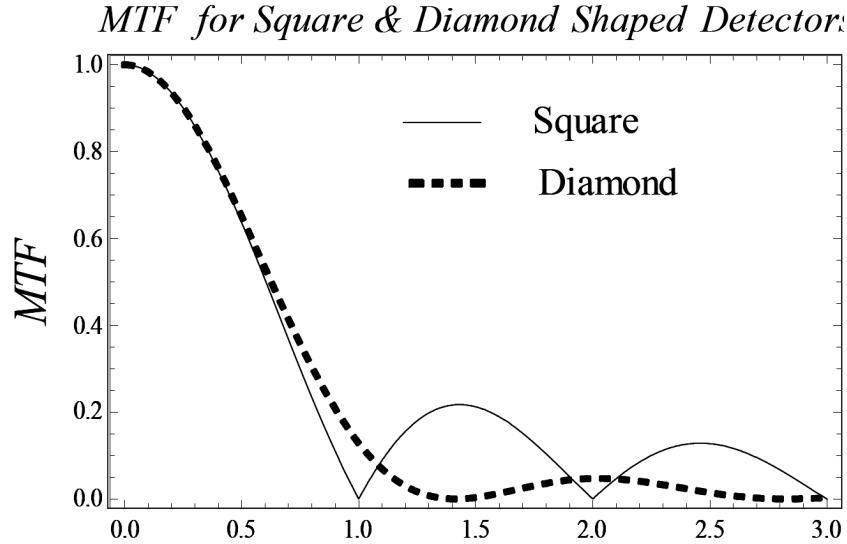


Figure 4.24 (a) Geometry of square and diamond-shaped detectors. (b) Detector MTFs of square and diamond-shaped detectors.

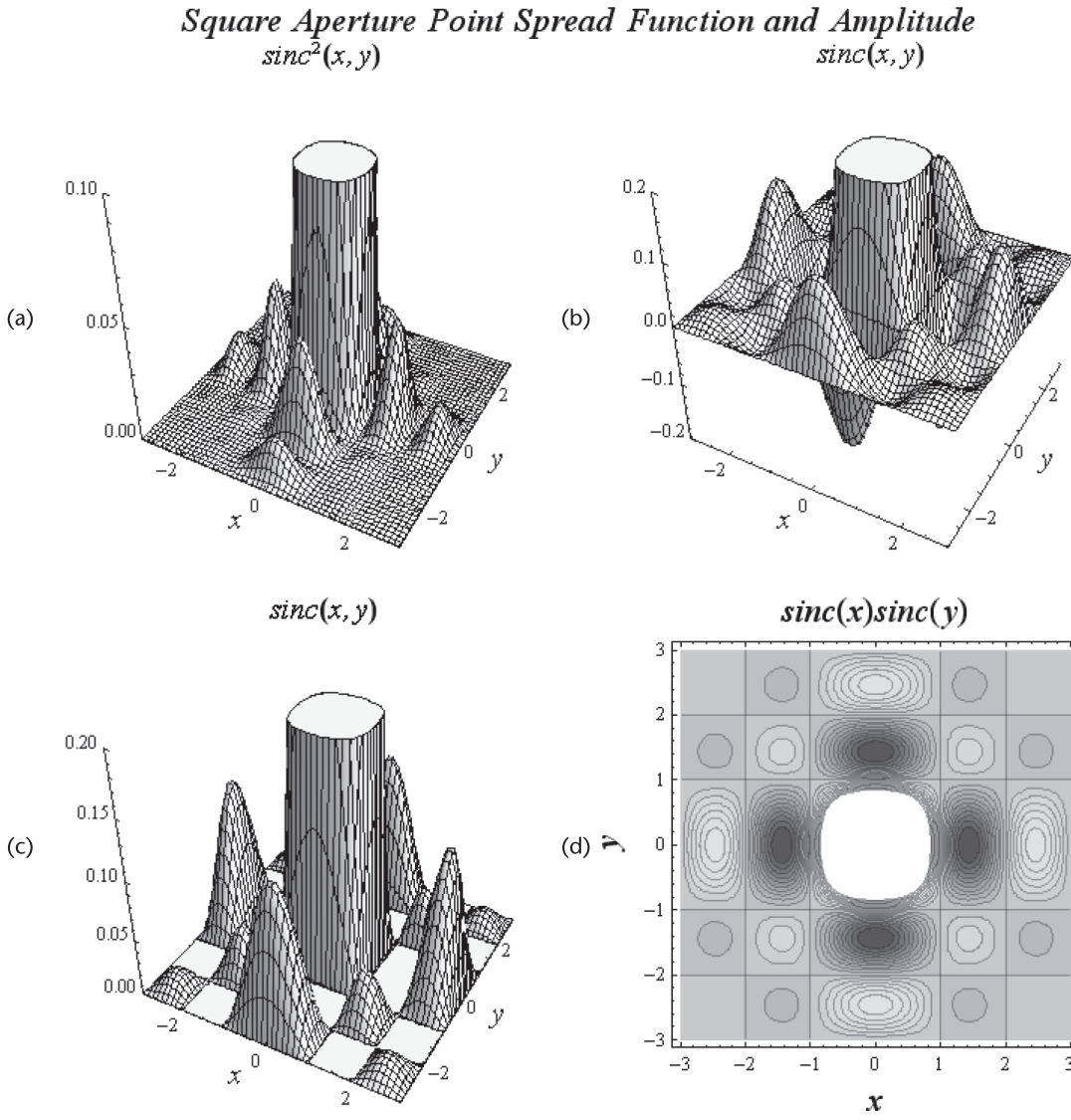
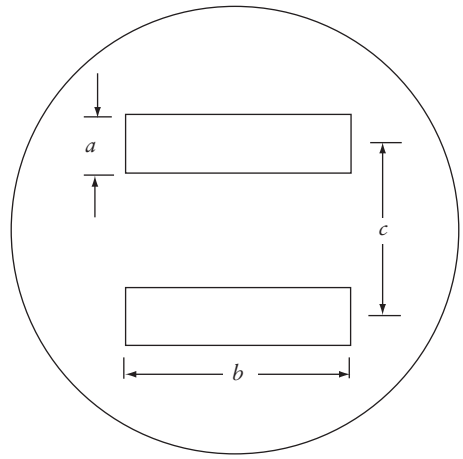


Figure 4.25 (a) Point spread function for square aperture. (b) Amplitude of optical disturbance is positive and negative. (c) For clarity, regions where the amplitude is negative have been set equal to zero. (d) Contour plot of amplitude.

a scanning system that utilized just the MTF in the horizontal and vertical directions would favor the diamond-shaped detector. Whether or not the diamond-shaped detector does perform better than the square detector in a scanning system is an open question.

4.10 Exercises

4.1 Assuming the Fraunhofer diffraction region, find the irradiance distribution of the double slit shown, given that it is illuminated by coherent unit amplitude, normally incident plane waves.



4.2 Find the Fresnel diffraction irradiance pattern caused by an aperture

$$p(x_a, y_a) = circ\left(\sqrt{x_a^2 + y_a^2}\,/\,a\right)$$

Assume a coherent, unit amplitude, normally incident plane wave with $\lambda = \lambda_0$.

4.3 A slit is illuminated by light of wavelength 0.5 μm. A diffraction pattern is formed on a screen 50 cm away from the aperture where the distance between the first and third minima is 2 mm. Determine the width of the slit.

4.4 A circular aperture 5 mm in diameter is illuminated by light of wavelength 0.5 μm. How far away must a screen be placed from the aperture to give the Fraunhofer diffraction pattern?

4.5 The spectral width of an argon laser is $\Delta\lambda$ = 0.2 μm and its wavelength is 0.488 μm. Compute the coherence length of the laser light.

4.6 An incoherent generalized optical system has a broadband (λ = 8 to 12 μm) infrared filter, an effective focal length of 20 cm, and a circular pupil (limiting aperture) of 10 cm.
a) Determine the diffraction-limited MTF of the system.
b) Determine the psf of the system.

4.7 Referring to Figure 4.16, show that

$$A_{OL}(s) = 2\left(\frac{D}{2}\right)^2\left[cos^{-1}\left(\frac{s}{D}\right) - \frac{s}{D}\sqrt{1-\left(\frac{s}{D}\right)^2}\right] \quad \text{for } s \le D$$

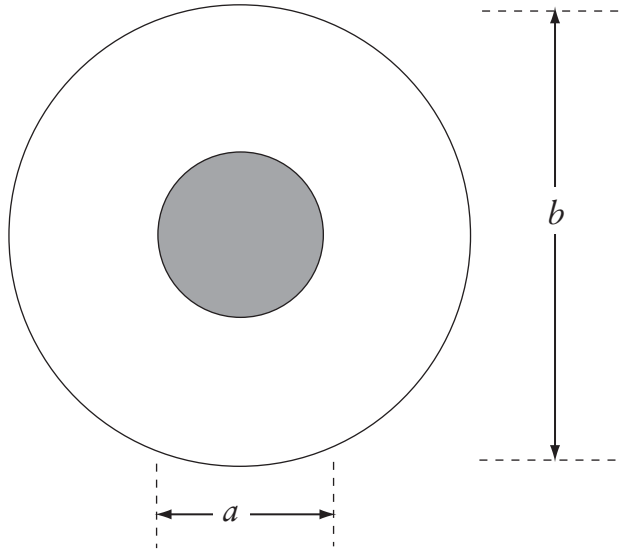
4.8 Show that the equations of (4.47) are correct.

4.9 Confirm the correctness of (4.63a) and (4.63b).

4.10 Numerically and analytically calculate the incoherent MTF of the following:
a) The elliptical entrance pupil of Figure 4.19

b) The distributed aperture of Figure 4.21
c) The racetrack aperture of Figure 4.23

4.11 Given the following figure:



a) Choose values for a and b and then calculate the diffraction MTF for the centrally obscured aperture shown.
b) Compare the MTF calculated above with the analytic result given in [8]

$$\mathrm{MTF}_{diff} = \frac{A + B + C}{1 - R_{obs}^2}$$

where

$$R_{obs} \equiv \frac{a}{b}$$

Define parameters X, Y and θ by the following relationships:

$$X \equiv \frac{f_x}{f_{oco}}, Y \equiv \frac{X}{R_{obs}}; \ \theta \equiv cos^{-1}\left(\frac{1 + R_{obs}^2 - 4X^2}{2R_{obs}}\right)$$

The variables A, B and C are:

$$A \equiv \begin{cases} \frac{2}{\pi}\left[cos^{-1}(X) - X\sqrt{1 - X^2}\right] & \text{when } 0 \le X \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$B \equiv \begin{cases} \dfrac{2R_{obs}^2}{\pi}\left[\cos{-1}(Y) - Y\sqrt{1-Y^2}\right] & \text{when } 0 \le Y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$C \equiv \begin{cases} -2R_{obs}^2 & \text{when } 0 < X \le \dfrac{1-R_{obs}}{2} \\ 0 & \text{when } X \ge \dfrac{1+R_{obs}}{2} \\ \dfrac{2R_{obs}}{\pi}\sin\theta + \dfrac{1+R_{obs}^2}{\pi}\theta - 2\dfrac{(1-2R_{obs})}{\pi}tan^{-1}\left[\left(\dfrac{1+R_{obs}}{1-R_{obs}}\right)\tan\left(\dfrac{\theta}{2}\right)\right] - 2R_{obs}^2 & \text{when } \dfrac{1-R_{obs}}{2} \le X \le \dfrac{1+R_{obs}}{2} \end{cases}$$

References

[1] Friedman, M., and J. Vizgaitis, "Calculating incoherent diffraction MTF," *Infrared Imaging Systems: Design, Analysis, Modeling and Testing XIX*, Proc. Of SPIE, Vol. 6941, 2008.

[2] Friedman, M., and M. Harwell, *Analog and Analytical Method for Determining Detector MTF*, United States Patent Number 5,075,883, 1991.

[3] Gillespie, C., *Dictionary of Scientific Biography*, New York City, NY: Scribner, 1972.

[4] Gaskill, J. D., *Linear Systems, Fourier Transforms and Optics*, Hoboken, NJ: Wiley, 1978.

[5] Goodman, J. W., *Introduction to Fourier Optics*, San Francisco, CA: McGraw-Hill, 2004.

[6] Guenther, R., *Modern Optics*, Hoboken, NJ: Wiley, 1990.

[7] Hecht, E., and A. Zajac, *Optics, Boston*, MA: Addison-Wesley, 1979.

[8] Holst, G., *Electro-Optical Imaging System Performance*, Bellingham, WA: 2008.

[9] Iizuka, K., *Engineering Optics*, New York City, NY: Springer-Verlag, 1983.

[10] Jenkins, F. A., and H. E. White, *Fundamentals of Optics*, New York City, NY: McGraw, 1976.

[11] Klein, M. V., *Optics*, Hoboken, NJ: Wiley, 1970.

[12] Nussbaum, A., and R. A. Phillips, *Contemporary Optics for Scientists and Engineers,* Upper Saddle River, NJ: Prentice-Hall, 1976.

[13] Pedrotti, F. L., and L. S. Pedrotti, *Introduction to Optics*, Hoboken, NJ: Prentice Hall, 1993.

[14] Saleh, B. E. A., and M. C. Teich, *Fundamentals of Photonics*, Hoboken, NJ: Wiley 1991.

[15] Williams, T., *Biographical Dictionary of Scientists*, Hoboken, NJ: Wiley, 1982.

[16] Yu, F. T., and I. C. Khoo, *Principles of Optical Engineering*, Hoboken, NJ: Wiley, 1990.

[17] Yu, F. T., *Optical Information Processing*, Hoboken, NJ: Wiley, 1983.

[18] Rutten, H., and M. van Venrooij, *Telescope Optics Evaluation and Design*, Richmond, VA: Willmann-Bell, 1988.

CHAPTER 5

Sources of Radiation

Now that the necessary mathematical, linear systems, and diffraction analytical tools have been covered, we turn our attention to the analysis of the system shown in Figure 5.1. The system comprises the radiation sources (targets and backgrounds), the atmosphere, and the sensor. The sensor includes optics, scanner (if required), detectors, electronics, display, and human perception. If an automated target processor is used, the display and human perception are not required. We return to this diagram frequently as a component road map to our analysis. The first component in the system is the objects that are viewed by the sensor. Objects are defined [1] as "figures seen through or imaged by an optical system." Objects, as seen by sensors, include targets and backgrounds. The objects of interest are targets, where backgrounds and clutter are those objects that reduce the system capacity to obtain accurate target information.

The goal of an observer is to distinguish the target from the background. Therefore, the sensor response to the target is frequently considered a signal and the sensor response to the background is considered a form of noise. Targets include people, buildings, ground vehicles, ships, aircraft, missile hardbodies, and missile plumes. The background and clutter include all the other radiation signals that enter the sensor aperture including signals from objects near the target, terrain or space behind the target, and the atmosphere. Atmospheric backgrounds can include clouds or other radiation variances that reduce the ability to discriminate the target signal. The atmospheric characteristics that affect sensor analysis are addressed in Chapter 6. Finally, clutter describes the background objects that are similar to the target in size and shape. Clutter makes the search process and discrimination of targets more difficult.

This chapter is organized to satisfy two goals: (1) to present the necessary background and material to characterize both infrared and EO signals originating from a target and background, and (2) to present the background and material to spatially describe a target. These signal and spatial characteristics can be combined with the two basic imaging system characteristics of sensitivity and resolution in order to determine the performance of an imaging system against a particular target and background.

Both infrared and EO signal characterization require a knowledge of radiometry, so this chapter begins with a radiometry/photometry section. Infrared signals are often characterized by a *delta T*, or a differential temperature between the target and the background. Even though this sounds like a thermal quantity, it is really a measure of optical power that is related to an equivalent blackbody differential temperature. The fundamentals of blackbody radiation are presented in the infrared signal section. EO signals are described by a target-to-background contrast in the

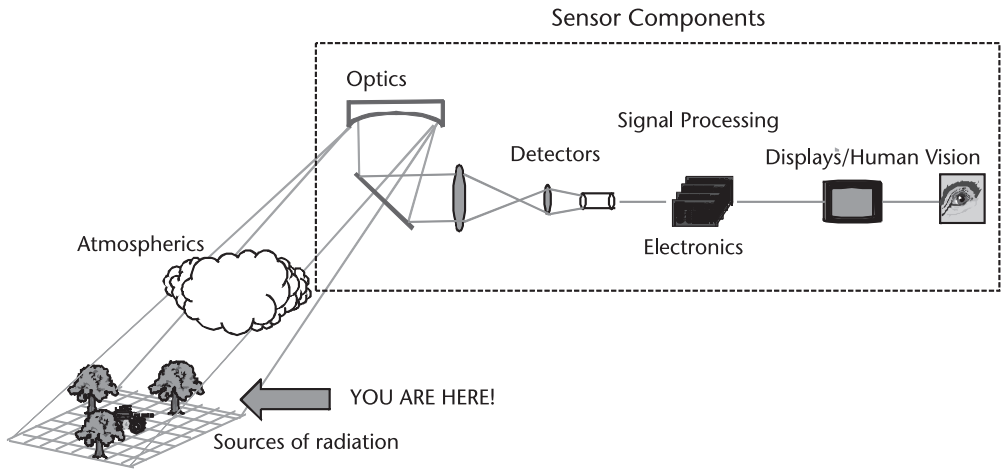


Figure 5.1 Targets and backgrounds are the first component in the overall system.

EO sensor band. Finally, the spatial characteristics of both infrared and EO systems are described in the last section along with the bar-target representation of targets.

5.1 Radiometry and Photometry

Sensor design and analysis revolve around some operational requirements that include a target under some background conditions. A typical requirement for a military infrared imaging system might be:

> The sensor shall provide a 90% recognition probability of a T-62 tank at a range of 2 km under mid-latitude summer weather conditions.

In a different application, an EO system may be designed around the recognition of people at an airport for security reasons. Both of these examples give a system requirement that is based on targets and must be translated into parameters that are meaningful for the sensor designer. The first step to understanding targets and backgrounds is to be fluent in radiometry.

Four photonic actions can occur at an object's surface. The quantities that characterize these actions are absorptance (α), reflectivity (ρ), transmissivity (τ), and emissivity (ε). The definitions of the first three quantities involve the ratio of absorbed, reflected, or transmitted energy to that of incident energy. Conservation of energy [2] dictates that

$$\alpha(\lambda) + \rho(\lambda) + \tau(\lambda) = 1 \tag{5.1}$$

Note that (5.1) is a function of wavelength. The majority of materials are not transmissive in the infrared wavelengths. A few exceptions are zinc selenide, sapphire, spinel, and germanium. Most of the materials that are transmissive in the EO wavelengths such as glass, plexiglass, plastic, and so forth are not transmissive in the infrared. For materials that are opaque ($\tau(\lambda) = 0$):

$$\alpha(\lambda) + \rho(\lambda) = 1 \tag{5.2}$$

In the EO wavelengths, external light (light from sources other than the target such as the sun, stars, and moon) provides illumination of the target. Here, highly reflective wavelengths (colors) have little absorption as seen in (5.2). In the infrared wavelength, the target provides most of the radiant signal seen by the sensor through emission while there is low reflectivity. In fact, for a target at thermal equilibrium (neither gaining nor losing heat), the absorptance and emissivity are identical:

$$\alpha(\lambda) = \varepsilon(\lambda) \tag{5.3}$$

Using (5.3) in (5.2),

$$\varepsilon(\lambda) + \rho(\lambda) = 1 \tag{5.4}$$

where it is now easy to see that a highly emissive infrared target has a low reflectivity.

In the infrared, the radiant signals come primarily from the target and background emission of blackbody radiation [3]. All bodies above the temperature of absolute zero (0K) emit electromagnetic waves. The heat in the body causes molecular vibrations, which, in turn, cause electron vibrations. These electrical vibrations provide the electromagnetic coupling to produce emission. A heated body placed in a vacuum chamber with total darkness (no impinging radiation) eventually loses all of its heat to the process of blackbody radiation. The amount of radiation that is emitted depends heavily on the target temperature and the surface characteristics (emissivity) of the target. A hotter target emits more radiation than a cooler target and a rough-surfaced target emits more radiation than a smooth target. These characteristics are also true for background objects and are discussed in more detail later in the chapter.

Reflected radiation in the infrared is usually small unless the radiation source is extremely bright as in the case of solar radiation. As a general rule, the reflection of solar radiation in the longwave (8 to 12 μm) is much less than reflections in the midwave (3 to 5 μm). Midwave sensors can view a significant amount of reflected solar radiation, so both reflected and emitted radiation must be considered for daytime scenarios.

With an EO sensor, the majority of the useful light in the bandwidth of detector responsivity is provided by the target and background reflectivities and some external (other-than-target) source. Most cases involve the sun as this illuminating source. Other sources include moonlight, starlight, and artificial lighting. Not only is the emission spectrum of the illuminating source important, but the reflectivity of the target is of utmost importance. This is the opposite of the longwave infrared sensor case. Here, the target radiation seen by the sensor is the illuminating spectrum of the source that is modified by the target reflectivity spectrum. The important contributors to the resulting contrast are the target reflectivity, background reflectivity, and illuminating spectrum.

Figure 5.2(a) shows a tank and a boat viewed with a longwave infrared sensor, and Figure 5.2(b) shows the same tank and boat viewed with a midwave infrared



Figure 5.2 Tank and boat images: (a) longwave, (b) midwave, (c) shortwave, and (d) visible. (Images provided by Ms. Stevie Smith, U.S. Army Night Vision and Electronics Sensors Directorate.)

sensor. The lighter shade of the tank depicts a larger optical radiation signal than the background and, therefore, a higher temperature or surface emission than the objects around the tank. One can quickly determine that a high infrared target-to-background contrast requires either high-temperature differential, emissivity differential, or a combination of both. Terrestrial targets are normally at temperatures (27°C) that contain high radiant power around wavelengths of 10 μm and are best viewed by longwave sensors. Rocket plumes, jet exhausts, and projectiles that are around 600° to 1,000°C give high radiant powers that are around 4 μm, suggesting that they are best viewed by the midwave infrared sensor. Hotter objects (the filament of a light bulb) can even be seen in the EO wavelengths.

The images shown in Figures 5.2(a) and (b) were taken during the day. The solar "loading" of the target occurs when the sun increases the temperature of parts of the tank or boat. The images at night do not have as much contrast since the variations in temperature are smaller. Also, the midwave signature during the daytime includes solar reflections. Figure 5.2(c) shows the same tank and boat as seen with a shortwave sensor, and Figure 5.2(d) shows the targets viewed with a visible sensor. Notice that the warm tracks on the tank are not resolved (here, they are in the shade) and that the images are constructed only by reflected energy from the sun.

Finally, in both EO and infrared systems, a very important parameter is the target size. Larger targets are obviously easier to detect, recognize, and identify. Smaller targets are difficult to resolve. The performance of a particular sensor in a scenario depends on the target size and target to background contrast. The transition of the operational sensor requirement to meaningful target and background parameters includes the representation of the radiant power leaving the target and background (i.e., contrast) as a function of space (i.e., target size). Once a target-to-background function is written in terms of the radiant power leaving that target as a function of space, the function can be transformed to the spatial frequency domain to determine which target and background spatial characteristics traverse through the sensor.

Radiometric Units

Radiometric concepts are necessary for a quantitative understanding of flux (power) transfer through a sensor. These concepts allow the analyst to determine how much energy is collected on the sensor detector surface. The amount of electromagnetic flux from both the target and the background contributes to the overall signal-to-noise ratio (SNR) of an image. Two important assumptions are made in radiometric analysis: (1) Sources are incoherent such that interference can be neglected, and (2) diffraction is ignored so that light is assumed to travel in straight lines.

The *solid angle* is a parameter that must be discussed because many radiometric units are given with respect to solid angles. A solid angle has units of steradians (sr) and is the angle subtended at the center of a sphere by an area on the surface of the sphere. The numerical value of a solid angle is a given area on the sphere surface divided by the square of the sphere radius. Consider the sphere shown in Figure 5.3. The solid angle is given by

$$\Omega = A/R^2 \tag{5.5}$$

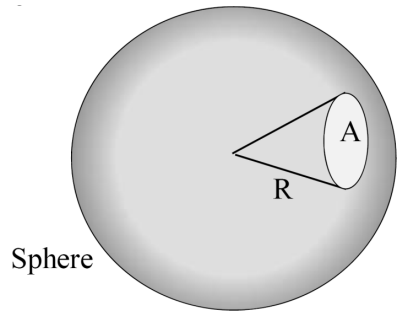


Figure 5.3 Depiction of a steradian.

where A is obtained from the curved surface of the sphere. The maximum number of steradians on a spherical surface is $4\pi R^2/R^2 = 4\pi$ sr. Note that an angle of radians multiplied by an angle of radians does not give steradians. That is, 2π radians times 2π radians gives $4\pi^2$ square radians, a factor of π off from the actual number of steradians in a sphere. Finally, if R is much larger than any dimension of the area, A, then the area on the sphere can be assumed to be flat. Given any flat shape imaginable and a distance to the shape, a steradian can be calculated for the corresponding solid angle using (5.5).

The units of radiometry [4] are shown in Table 5.1, where the basic unit of energy is given in terms of the joule. The amount of energy per unit time is flux, or power. Intensity is the amount of power a source delivers per unit solid angle. Exitance and irradiance both have the same units of power-per-unit area; however, exitance addresses power leaving a surface and irradiance addresses the amount of power impinging on a surface. One interesting note is that exitance is sometimes called *emittance*, depending on the community. Exitance really describes the total power per unit area leaving an object including the emitted power and the reflected power. Each of the units of Table 5.1 is used, whether directly or indirectly, during a source-to-detector radiometry calculation. Many times the partial differentials yield to simpler calculations if the amount of power leaving or impinging on a surface is constant across the surface.

Table 5.1 Radiometric Units

Symbol	*Quantity*	*Definition*	*Units*
Q	Energy	Fundamental	Joule
Φ	Flux (Power)	$\frac{\partial Q}{\partial t}$	J/s = W
I	Intensity	$\frac{\partial Q}{\partial t \partial \Omega} = \frac{\partial \Phi}{\partial \Omega}$	W/sr
M	Exitance or Emittance	$\frac{\partial^2 Q}{\partial t \partial A} = \frac{\partial \Phi}{\partial A}$	W/cm^2
E	Irradiance	$\frac{\partial^2 Q}{\partial t \partial A} = \frac{\partial \Phi}{\partial A}$	W/cm^2
L	Radiance	$\frac{\partial^3 Q}{\partial t \partial \Omega \partial A} = \frac{\partial \Phi}{\partial A \partial \Omega} = \frac{\partial I}{\partial A}$	W/(cm^2-sr)

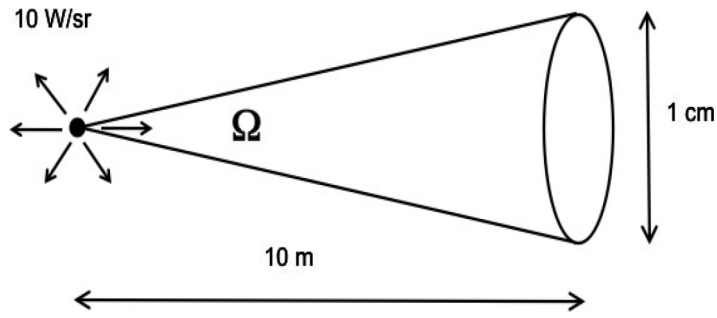


Figure 5.4 Point-source intensity example.

Example 5.1
An example of the simplicity of radiometric calculations for uniform sources is illustrated by the following. A 1-mm × 1-mm square emits 1 μW of light uniformly over the area. The corresponding emittance is 1 μW divided by the small area, giving 100 μW/cm^2.

Example 5.2
Frequently, point-source radiation is specified in terms of its intensity. The point source shown in Figure 5.4 provides a uniform 10 W/sr intensity over an entire sphere. The power entering a 1-cm aperture located 10 m from the point source is desired.

The solution is given by first determining the solid angle of the aperture, Ω. The solid angle is given by the area of the aperture divided by the square of the range (distance from the source to the aperture), or $\Omega = [\pi(0.005)^2]/[10^2]$ giving 7.85×10^{-7} sr. Because the point source provides uniform illumination over all angles, the amount of power through the aperture is 10 W/sr (7.85×10^{-7} sr) 7.85 μW.

Radiance is given as the source power per unit area per unit source angle. Radiance is a surface emission quantity that is independent of the source area or solid angle. This is a quality that becomes apparent in the blackbody section.

Example 5.3
A 1-mm radius sphere provides 1 W of power uniformly in all directions. Find the radiance of the sphere. Assuming a uniform radiance over the sphere, the solution is given by normalizing the power by the sphere surface area $4\pi(0.1)^2$ square cm and the spherical solid angle 4π steradians to give $L = 0.63$ W/(cm^2-sr).

One particular surface quality applies to a majority of natural and human-made surfaces. The Lambertian surface includes most rough (rough on the order of a wavelength) surfaces such as paints, cloths, stones, and wood that is not shiny. Specular, or shiny and mirrored, surfaces are usually not Lambertian. The Lambertian surface follows a decrease in intensity with angle

$$I_\theta = I_n \cos\theta \quad \text{[W/sr]} \tag{5.6}$$

as shown in Figure 5.5.

The value I_n represents the intensity normal to the surface and the angle θ is taken from the normal direction. The corresponding radiance perceived by a viewer of a tilted Lambertian surface is constant with angle

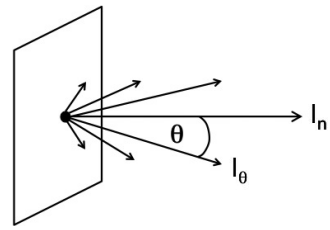


Figure 5.5 Lambertian surface.

$$L_\theta = \frac{\partial I}{\partial A} = \frac{I_\theta}{dS_\theta} = \frac{I_n cos\theta}{dA cos\theta} = \frac{I_n}{dA} \quad [\text{W}/(\text{cm}^2\text{-sr})] \tag{5.7}$$

The decrease in intensity from the tilt is compensated by the increase in perceived area by the viewer so that the surface appears to remain at a constant brightness. The Lambertian surface is a perfect diffuser that gives a viewer a constant radiance independent of viewing angle. Also, a Lambertian [5] surface gives a radiance that is directly proportional to its emittance:

$$L = \frac{M}{\pi} \quad [\text{W}/(\text{cm}^2\text{-sr})] \tag{5.8}$$

This characteristic proves to be convenient in the manipulation of thermal sources because their radiation characteristics are defined by emittance.

Photometric Units

Photometric units are very similar to radiometric units; however, the radiation is weighted to match the response of the human eye. A luminous efficacy $K(\lambda)$ describes the weighting of the radiation to provide an equivalent photometric response. The photometric, or luminous, flux is found by

$$\Phi_v(\lambda) = K(\lambda)\Phi(\lambda) \tag{5.9}$$

where the v subscript denotes the visual photometric response and K is in units of lumens per watt. A photopic [6] efficiency is the normalized efficacy

$$V(\lambda) = \frac{K(\lambda)}{K_{\max}} \tag{5.10}$$

where $K_{\max}$ is 673 L/W. The photopic response described above is formed primarily by the contributions from cones within the human retina. Cone receptors resemble high-resolution, slow-speed film. The night, or scotopic, response of the human retina is found using $K'(\lambda)$, which was derived by the rod receptors of the human retina. The corresponding $V'(\lambda)$ is given with a normalization $K'_{\max}$ of 1,725 L/W. Rod receptors are more like low-resolution, high-speed film. The efficiencies of both photopic and scotopic visions are given in Figure 5.6.

Values for $K(\lambda)$ and $K'(\lambda)$ can be found a number of National Bureau of Standards (NBS) reports. Note that the photometric equivalents to the radiometric quantities of radiance, intensity, exitance, or irradiance are found by using the photopic or scotopic efficacy functions multiplied by the fundamental flux $\Phi(\lambda)$. These cor-

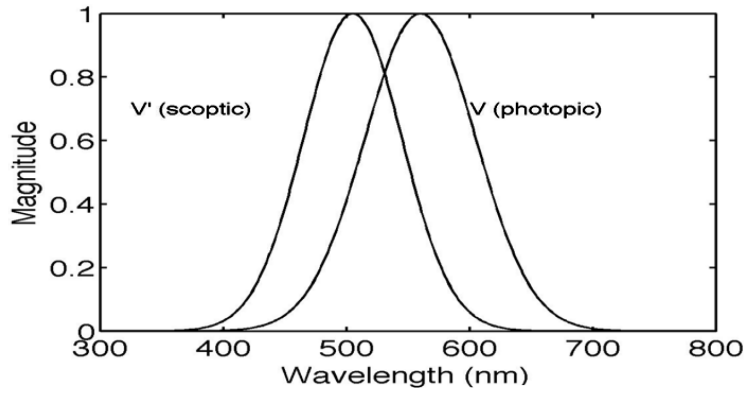


Figure 5.6 Photopic and scotopic efficiencies.

responding photometric quantities are described as luminance, luminous intensity, luminous exitance, and illuminance. All of these quantities can be written like those in Table 5.1 with the ν subscript, as shown in Table 5.2. Also, the NBS standard is now established as the *candela* [7] that corresponds to a luminous intensity of 1.0 L/sr. Many other photometric units have been defined, such as the Carcel unit, English sperm candle, and the Hefner unit, but are rarely used. Finally, it is only necessary to integrate the efficacy-power product of the photometric quantities from 380 to 780 nm because these are the limits of the human visual response.

5.2 Infrared Targets and Backgrounds

Infrared target and background characteristics in the 3- to 5-μm band and the 8- to 12-μm band are of primary interest. These wavelengths correspond to windows in Earth's atmosphere where light propagates with less severe attenuation. Even in these two windows, the attenuation can be significant. The majority of target characterization work has been performed in these bands.

Target characterization is an important part of the overall sensor analysis and design process. Sensor band selection for a scenario begins with the targets and backgrounds. A target-to-background difference in exitance (or emittance) must

Table 5.2 Photometric Units

Symbol	*Quantity*	*Definition*	*Units*
F_υ	Luminous flux (power)	$\Phi_\upsilon = \frac{\partial Q}{\partial t}$	Lumens (L)
I_υ	Luminous intensity	$\frac{\partial \Phi_\upsilon}{\partial \Omega}$	L/sr
M_υ	Luminous exitance	$\frac{\partial \Phi_\upsilon}{\partial A}$	L/m^2 or lux
E_υ	Illuminance	$\frac{\partial \Phi_\upsilon}{\partial A}$	L/m^2 or lux
L_υ	Luminance	$\frac{\partial \Phi_\upsilon}{\partial A \partial \Omega} = \frac{\partial I_\upsilon}{\partial A}$	L/(m^2-sr)

be present in the band of interest. Exitance is the area-normalized quantity that describes the flux leaving a surface. For infrared systems used during the day, solar reflections must be addressed and exitance must be considered as the sum of emitted and reflected flux. Many tactical systems that include infrared sensors also have EO systems. Because EO systems have higher resolution and more sensitivity than infrared systems during the day, the use of infrared systems is typically reserved for night. Therefore, the focus here is on night use and the emittance of targets and backgrounds.

Blackbody Radiation

In the infrared world, the absorption and the emission characteristics of targets and backgrounds are usually more significant than their reflection and transmission characteristics. This is especially true at night (the usual mode of infrared system operation), where solar reflection is not present. Daylight hours give solar reflections even in the infrared. However, many natural and human-made objects are low in infrared reflectance and high in emissivity (and absorptance). These objects tend to absorb solar energy with subsequent emission under conditions of thermal equilibrium.

Blackbody radiation is radiation that is released from bodies that are above the temperature of absolute zero (0 K). The finite temperature of the bodies and their associated heat content provide the energy necessary to emit electromagnetic, or photonic, waves. Two basic concepts hold true for blackbody radiation: (1) A higher temperature body corresponds to a larger emission of flux, and (2) a higher temperature body shifts the flux spectral distribution toward shorter wavelengths. These concepts are true regardless of the emissivity characteristics of the body (including targets and backgrounds).

The spectral distribution and magnitude of an object's radiation are primarily a function of the object's temperature and emissivity. Emissivity, $\varepsilon(\lambda)$, describes an object's emission characteristics where an emissivity of one, $\varepsilon(\lambda) = 1$, indicates a perfect blackbody radiator. Blackbody radiation is described by Planck's blackbody equation. The spectral emittance of an object is

$$M(\lambda, T) = \varepsilon(\lambda) \frac{c_1}{\lambda^5} \frac{1}{\left[e^{\frac{c_2}{\lambda T}} - 1\right]} \quad [\text{W}/(\text{cm}^2\text{-}\mu\text{m})] \tag{5.11}$$

where

λ = the wavelength in micrometers (μm)
T = absolute temperature in Kelvin (Kelvin = Celsius + 273)
$c_1 = 3.7418 \times 10^4$ W-μm^4/cm^2
$c_2 = 1.4388 \times 10^4$ μm-K.

The units show a distinction between the integration over wavelength and the spatial normalization of the emittance (i.e., cm^2-μm). Figure 5.7 shows the emission spectrum of a perfect blackbody at a number of temperatures. Note the dramatic increase in the overall emittance with temperature and the corresponding shift in the peak emission wavelength. The peak wavelength of emission for a given blackbody temperature is given by Wien's law:

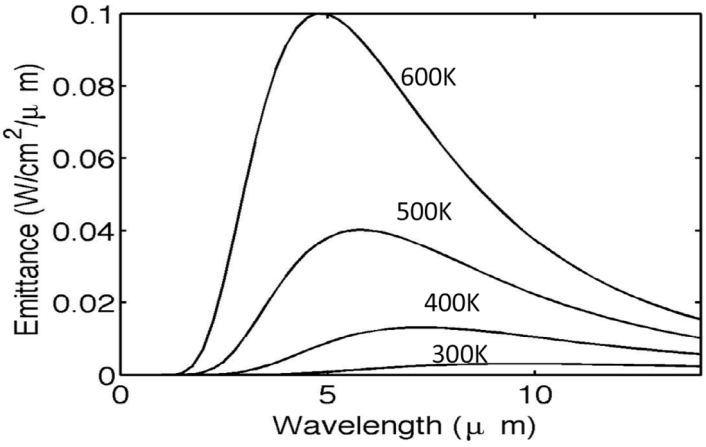


Figure 5.7 Planck's blackbody emission curves.

$$\lambda_{peak} = \frac{2897.8}{T} \quad [\mu m] \tag{5.12}$$

where the temperature is given in degrees kelvin. The total emittance within a spectral band is found by summing, or integrating, the emittance over the appropriate wavelengths:

$$M(T) = \int_{\lambda_2}^{\lambda_1} \varepsilon(\lambda) \frac{c_1}{\lambda^5} \frac{1}{[e^{\frac{c_2}{\lambda T}} - 1]} d\lambda \quad [W/cm^2] \tag{5.13}$$

This equation is usually integrated numerically or found using lookup tables. Alternatively, one can use the series solution given by [21]. Their solution requires few terms, yet is extremely accurate over a very broad range of wavelengths.

An object with a constant emissivity over all wavelengths is called a graybody. Blackbodies and graybodies exhibit an interesting property called the Stephan-Boltzmann Law. This law provides simple flux calculations when the emission over the entire spectrum is desired:

$$M(T) = \varepsilon \sigma T^4 \quad [W/cm^2] \tag{5.14}$$

The Stephan-Boltzmann constant σ is 5.67×10^{-12} W/(cm^2-K^4). The equation applies only to blackbody and graybody sources. Objects that emit as a function of wavelength, where they do not follow blackbody or graybody characteristics, are described as *spectral emitters*. The total or band-limited emittance of a spectral emitter and the band-limited emittance of blackbodies and graybodies must be evaluated using (5.13). Finally, blackbodies, graybodies, and the majority of spectral emitters can be considered Lambertian sources. The radiances of these sources are found simply by using (5.8).

Example 5.4

A flat, 500°C (773 K) blackbody source is considered Lambertian. Determine the peak emittance of the source along with the corresponding wavelength. Determine the emittance and radiance of the source in the 3- to 5-μm band. Finally, determine

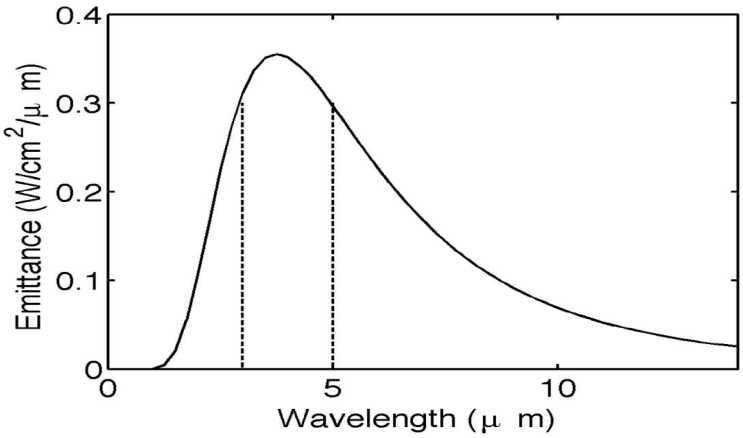


Figure 5.8 Midwave example with 500°C blackbody.

the total emittance and radiance of the source (over all wavelengths). The solution is given using Figure 5.8 as a guide.

The peak emission wavelength is determine to be 3.75 μm using Wien's law given by (5.12). The peak emission (at 3.75 μm) is calculated using (5.11), Planck's blackbody equation, to give 0.355 W/(cm^2-μm). In the 3- to 5-μm band, Planck's equation is integrated over these wavelengths, which gives an emittance of 0.67 W/cm^2. Assuming a Lambertian source, (5.8) gives a radiance of 0.213 W/cm^2-sr. The Stephan-Boltzmann law gives the total emittance of the source at 2.02 W/cm^2 and a total radiance of 0.64 W/(cm^2-sr).

Example 5.5

A 20°C (293K) graybody source is flat and is considered Lambertian. The emissivity is 0.5 over all wavelengths. Determine the emittance and radiance of the source in the 8- to 12-μm band. Also, determine the total emittance and radiance of the source (over all wavelengths). The solution can be seen by reviewing Figure 5.9.

The peak wavelength of emission is 9.89 μm, in the center of the 8- to 12-μm

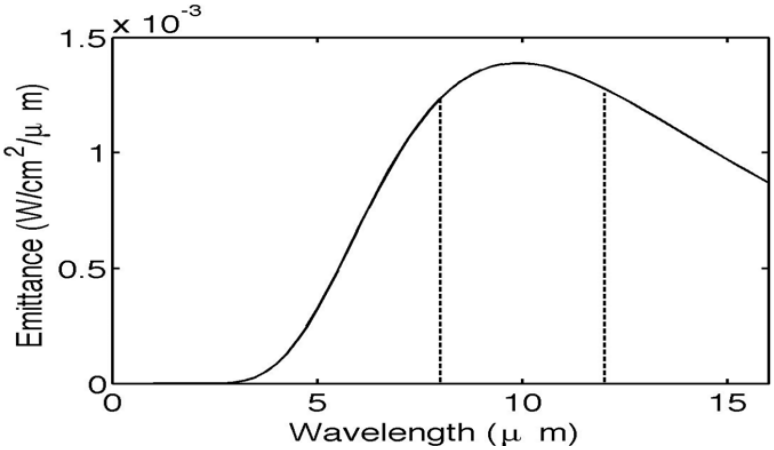


Figure 5.9 Longwave example with 20°C graybody.

band. The band emittance is calculated to be 5.3×10^{-3} W/cm2 and the band radiance is 1.7×10^{-3} W/cm^2-sr. The total emittance over all wavelengths is 21×10^{-3} W/cm^2 and the total radiance is 6.7×10^{-3} W/cm^2-sr.

The spectral emittance of a target or background is described by Planck's equation, (5.11). Note that the surface emittance of an object depends on object temperature, wavelength, and emissivity. Given that an object is viewed by a sensor with a limiting spectral bandwidth, the emittance of the object is integrated over the band to determine the object's emittance as seen by the sensor. This bandwidth integration must include atmospheric, sensor optics, and detector spectral responses. A common and incorrect analysis method includes a band-averaged target emittance propagated through a band-averaged atmospheric attenuation and a band-averaged sensor transmission and collected by a band-averaged detector response. This analysis method is used frequently by many analysts and gives a result that includes a level of error. FLIR92 and the NVL TV model are examples of analytical tools that use this technique, where their associated errors are determined to be acceptable. The errors decrease with object emissivities, atmospheric transmissions, and detector responses that do not vary much over the integration band. Smaller bands also give smaller errors.

Whether we use an exact solution or a band-averaged estimate, it is useful to band-average an object's emittance to get a feel for the differences in target-to-background emittance. Given a particular spectral band, the emittance of an object can be found through the integration of (5.13) using the object's temperature and emissivity. This integration is simple if the object is a blackbody (emissivity = 1) or a graybody (emissivity = constant). Lookup tables for the emittance of objects at different temperatures have been compiled and can be found in a number of the references. The emittance must be calculated numerically for objects that are considered spectral emitters.

Emissivity

Emissivity can be described in a number of ways, one of which is the spectral, hemispherical emissivity:

$$\varepsilon(\lambda) = \frac{M_{object}(\lambda)}{M_{BB}(\lambda)} \quad \text{[unitless]} \tag{5.15}$$

where the emission of an object is normalized by the emission of a blackbody source at the same temperature. Blackbody sources are readily available and widely used in the infrared community. They are typically cavities or flat surfaces that have been treated to provide an extremely high emissivity over all infrared wavelengths. Blackbodies are also used as references in the measurement of object-emission or sensor-response characteristics.

Different techniques are used to measure object emissivity. One of the more straightforward techniques involves the use of a blackbody positioned next to the object, where the blackbody temperature is set to match that of the object. The object temperature is measured with a high-resolution thermocouple in a manner that does not change the object's temperature. An infrared spectral radiometer collects flux from the object as a function of wavelength (i.e., a spectral scan) and

then collects flux from the blackbody source as a function of wavelength. Each flux measurement in the object spectral scan is divided by the corresponding flux measurement of the blackbody spectral scan to give the emissivity as a function of wavelength. Figure 5.10 shows the spectral emissivity for a number of objects.

Another representation of emissivity is that of total emissivity. This is a single emissivity value over all wavelengths. That is, the infinite integral of an object's exitance spectrum relative to that of a blackbody:

$$\varepsilon = \frac{\int_o^\infty M_{object}(\lambda)\, d\lambda}{\acute{o}T^4} \tag{5.16}$$

Note that the wavelength dependence of emissivity has been dropped. This type of emissivity is useful if the object is a graybody because the emission spectrum of the object can then be determined. Care should be used, however, when a single emissivity value is used to represent the emission of a spectral emitter over wide spectral bands where the actual emissivity would vary significantly. For example, the emissivity of olive-drab uniform material could be considered a graybody over an 8- to 10-μm spectral band with an emissivity value of around 0.96 to 0.97. This is a poor assumption for white paint over the 3- to 5-μm spectral band. The multiplication and integration of this emissivity curve by a 3- to 5-μm blackbody emission would give significantly different results than would a constant emissivity approximation. As a rule, most objects are closer to a graybody source in the 8- to 12-μm spectral band than in the 3- to 5-μm spectral band. Analysts find frequently that approximations in the 3- to 5-μm spectral band yield significant errors.

Some common examples for total emissivities [8] are shown in Table 5.3. Note that shiny metal objects are usually less emissive and more reflective, whereas rough metal objects are more emissive and less reflective. Also, while water and glass have smooth surfaces, their emissivities are high. Rough, or diffuse, objects can usually be considered Lambertian sources. Shiny objects are considered special cases of reflectors (sometimes called specular objects). Finally, a very interesting emissivity is that of human skin, which provides an outstanding infrared source. Humans are almost perfect radiators.

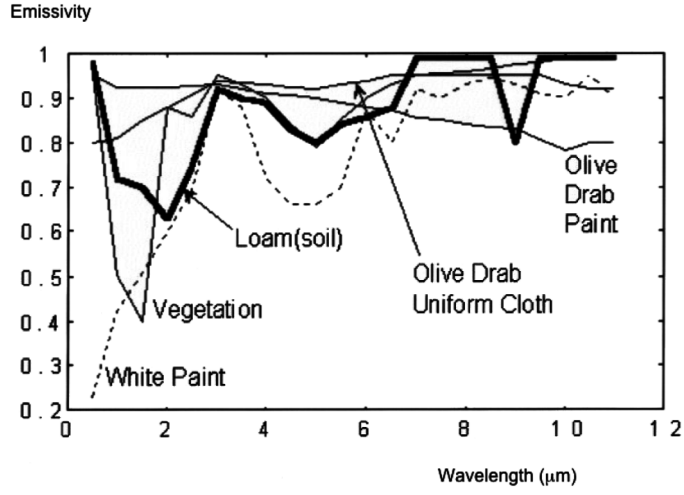


Figure 5.10 Spectral emissivity for various objects (targets and backgrounds).

Table 5.3 Total Emissivities

Material	*Emissivity*
Polished aluminum	0.05
Anodized aluminum	0.55
Polished brass	0.03
Oxidized brass	0.61
Polished gold	0.02
Polished steel	0.07
Oxidized steel	0.79
Tin plate	0.07
Brick	0.93
Graphite	0.98
Polished glass	0.94
Oil paint (average of 10 colors)	0.94
Paper	0.93
Sand	0.90
Human skin	0.98
Soil	0.92
Snow	0.98
Water	0.96
Oak wood (planed)	0.90

Equivalent Differential Temperature (Delta T)

Target-to-background contrast is described in the infrared with the equivalent differential temperature, or *delta T*. Although this quantity may appear to be a thermal quantity, it is really a radiometric quantity. It is the equivalent differential temperature (at some prescribed background temperature) of two blackbody sources, target and background, required to provide the same differential flux as that of the actual target and actual background. This equivalence is, of course, in the band of the sensor. Consider the blackbody source and infrared sensor shown in Figure 5.11. Using Figure 5.11 as a guide, the radiance of the target is given, and the source area

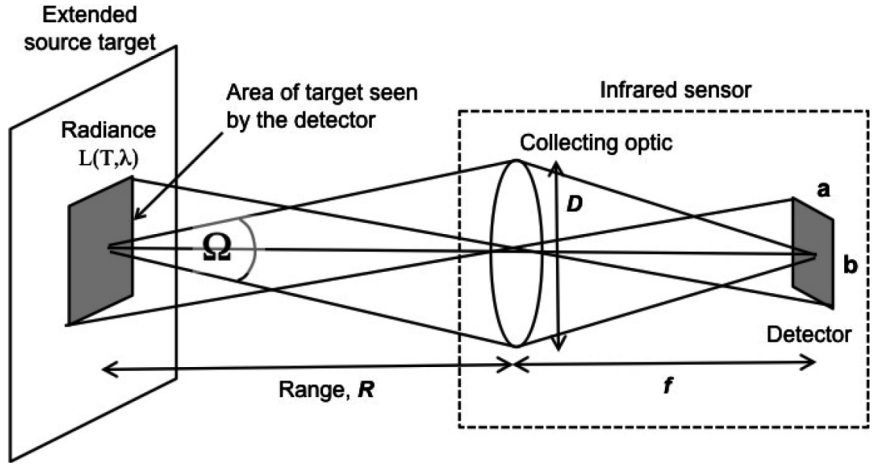


Figure 5.11 Differential temperature geometry.

seen by a single detector is

$$A_{src} = \frac{abR^2}{f^2} \quad [\text{cm}^2] \tag{5.17}$$

where a and b are the dimensions of the detectors in centimeters, f is the focal length of the sensor in centimeters, and R is the distance from the sensor to the target in centimeters. The solid angle of the sensor as seen from the source is given as

$$\Omega = \frac{\pi D^2}{4R^2} \quad [\text{sr}] \tag{5.18}$$

where D is the diameter of the entrance aperture in centimeters. The total target flux falling on the detector is given by

$$\Phi_{tgt} = \frac{abR^2}{f^2}\frac{\pi D^2}{4R^2}\int_{\lambda_1}^{\lambda_2} L_{tgt}(T,\lambda)\, d\lambda = \frac{\pi abd^2}{4f^2}\int_{\lambda_1}^{\lambda_2} L_{tgt}(T,\lambda)\, d\lambda \quad [\text{W}] \tag{5.19}$$

If the detector were receiving flux from the background, the amount of flux would be

$$\Phi_{bg} = \frac{\pi abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2} L_{bg}(T,\lambda)\, d\lambda \quad [\text{W}] \tag{5.20}$$

The differential flux between a detector viewing the target and a detector viewing the background is

$$\Delta\Phi = \frac{\pi abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2}\left[L_{tgt}(T,\lambda) - L_{bg}(T,\lambda)\right] d\lambda \quad [\text{W}] \tag{5.21}$$

This equation can be extended to give a differential voltage on the output of the detector from the difference in flux between the target and the background. Using the transmission of the optics $\tau_{optics}(\lambda)$ and the responsivity of the detector in V/W, $R(\lambda)$:

$$\Delta V = \frac{\pi abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2}\left[L_{tgt}(T,\lambda) - L_{bg}(T,\lambda)\right]\tau_{optics}(\lambda)R(\lambda)\, d\lambda \quad [\text{V}] \tag{5.22}$$

Now, given some background reference temperature T_{bg}, an equivalent blackbody ΔT can be calculated, which causes the same differential voltage:

$$\Delta V = \frac{\pi abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2}\left[L_{bb}\left(\Delta T + T_{bg},\lambda\right) - L_{bb}(T,\lambda)\right]\tau_{optics}(\lambda)R(\lambda)\, d\lambda \quad [\text{V}] \tag{5.23}$$

where $L_{bb}(T, \lambda)$is the radiance of a blackbody source. Equating the two differential voltages [of (5.22) and (5.23)] gives the solution for the delta T:

$$\begin{aligned}&\int_{\lambda_1}^{\lambda_2}\left[L_{tgt}(T,\lambda) - L_{bg}(T,\lambda)\right]\tau_{optics}(\lambda)R(\lambda)\, d\lambda = \\ &\int_{\lambda_1}^{\lambda_2}\left[L_{bb}(\Delta T + T_{bg},\lambda) - L_{bb}(T_{bg},\lambda)\right]\tau_{optics}(\lambda)R(\lambda)\, d\lambda\end{aligned} \tag{5.24}$$

Note that the target radiance, background radiance, optical transmission, and detector responsivity must be known. Also, there is no closed-form solution for the delta *T*, so the solution is usually obtained using numerical methods

Apparent Differential Temperature (Apparent Delta T)

A more important quantity than delta *T* is the *apparent delta T*, or the equivalent blackbody differential temperature that would cause the same differential voltage as that of the target and background *as seen through some atmospheric path* (Figure 5.12). A simple way to describe the apparent delta *T* is that a differential voltage is seen on the output of a detector corresponding to the differences in flux of the target and the background seen through some atmospheric path. The apparent delta *T* would be the differential temperature (with respect to some background temperature) of two nearby (through negligible atmospheric path) blackbody sources that produce the same differential voltage. A number of methods are used to calculate apparent delta *T*.

The following techniques are used to determine the apparent target-to-background differential temperature at the entrance aperture of an infrared sensor. All of these techniques have been used in analyses. We assume that the apparent temperature corresponds to the actual target-to-background radiance reduced by the atmospheric transmission between the target and the infrared sensor.

The apparent temperature is an equivalent blackbody target temperature differential (above an ambient background . . . 300K) that creates an identical differential voltage as that of the target and background through the atmosphere. The real apparent differential temperature is a function of background flux, target flux, and atmospheric transmission, where each is expressed as a spectral quantity. The following calculations are approximations of apparent temperature. Note that the quantity τ_{atm} is a band-averaged atmospheric transmission quantity equal to

$$t_{atm} = \frac{1}{\Delta\lambda}\int_{\lambda_1}^{\lambda_2} \tau_{atm}(\lambda)\, d\lambda \qquad \text{[unitless]} \quad (5.25)$$

This quantity is used in some of the band-averaged approximations below. Each of these approximations is a function of range, however, the range dependence is addressed in each description.

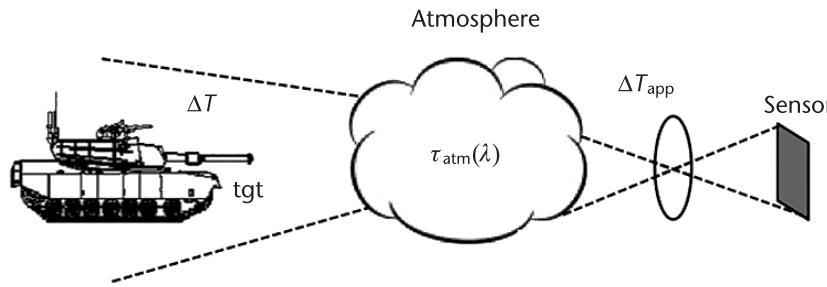


Figure 5.12 Apparent delta T.

Technique 1: Temperature–Broadband Beer's Law Product

The first technique is that of an apparent temperature determined by the product of the target-to-background temperature and the atmospheric broadband Beer's law transmission. That is:

$$\Delta T_{app} = \Delta T_{tgt} \tau_{1\,km}^{R} \quad [\mathrm{K}] \tag{5.26}$$

The broadband transmission shown is determined for a 1-km path length and R is the target-to-sensor range in kilometers. The 1-km path length transmission can be found by averaging the spectral transmission for 1 km. Beer's law, discussed in Chapter 6, is not generally accurate for the broadband transmission of light. This technique is one of the more common approximations for calculating apparent temperature (especially in the longwave).

Technique 2: Temperature–Broadband Transmission as a Function of Range

The second technique is similar to the first technique in that the apparent temperature is determined by a target temperature–broadband transmission product. The difference here is that a broadband Beer's law is not assumed, and the broadband transmission is determined using an atmospheric transmission program as a function of range. That is, an atmospheric program, such as MODTRAN, is run for a number of desired ranges, and the spectral transmission is averaged for each range to give a broadband transmission as a function of range. From the MODTRAN output, a table is formed with range and its corresponding broadband transmission. If the transmission for some range is needed that is not listed in the table, the transmission is interpolated from the tabular data. The apparent temperature is then determined by

$$\Delta T_{app} = \Delta T_{tgt} \tau(R) \quad [\mathrm{K}] \tag{5.27}$$

This technique is the method used in the acquisition program ACQUIRE, the most common sensor/scenario engagement calculation. Note that this technique eliminates the errors associated with a broadband Beer's law assumption, however; the errors associated with a nonspectral flux calculation are still present.

Technique 3: Flux–Broadband Beer's Law Product

Another technique is that of broadband radiance multiplied by the Beer's law broadband transmission approximation to give the apparent radiance. This is a flux-based calculation, but still a gross approximation because of the broadband radiance and Beer's law associated errors:

$$\Delta L_{app} = \Delta L_{tgt} \tau_{1\,km}^{R} \quad [\mathrm{W/cm^2\text{-}sr}] \tag{5.28}$$

The procedure for this calculation is described as follows: The broadband differential target-to-background radiance is calculated for some target radiance and background radiance. Once the differential radiance is computed, this target-to-background differential radiance is multiplied by the Beer's law broadband trans-

mission to obtain an apparent differential radiance. Using this differential radiance, a blackbody source temperature and a blackbody background temperature (ambient background) are estimated that together provide a band-integrated differential radiance equal to the apparent radiance. This is a trial-and-error or lookup table procedure. This technique can also be extended with a broadband transmission that is calculated for a number of ranges (as in technique 2).

Technique 4: Flux–Temperature Differential

Lloyd [9] derives an apparent differential temperature based on the differential radiance with respect to the background temperature. The target-to-background signal is propagated through the atmosphere to give an equivalent differential temperature:

$$\Delta T_{app} = \frac{\Delta V}{\partial V / \partial T} = \frac{\int_o^\infty \left[L_{tgt}(\lambda) - L_{by}(\lambda) \right] \tau_{atm}(\lambda) \tau_{optics}(\lambda) R(\lambda)\, d\lambda}{\int_o^\infty \frac{\partial L_{bb}(T_{by},\lambda)}{\partial L_{bg}} \tau_{atm}(\lambda) \tau_{optics}(\lambda) R(\lambda)\, d\lambda} \quad \text{[K]} \qquad (5.29)$$

This expression requires the responsivity $R(\lambda)$ of the detector and the transmission of the optical system $\tau_{optics}(\lambda)$. The term ΔV is the differential voltage on the output of the detector caused by the target-to-background differential flux.

Technique 5: Exact Transcendental Solution

The exact solution can be written in a manner that equates the differential voltage on the output of the detector corresponding to the target-to-background differential flux to that of close (atmospheric degradation is negligible) blackbody sources:

$$\begin{aligned} &\int_0^\infty \left[L_{tgt}(\lambda) - L_{by}(\lambda) \right] \tau_{atm}(\lambda) \tau_{optics}(\lambda) R(\lambda)\, d\lambda \\ &\int_0^\infty \left[L_{bb}(\Delta T_{app} + T_{bg}, \lambda) - L_{bb}(T_{bg}, \lambda) \right] \tau_{optics}(\lambda) R(\lambda)\, d\lambda \end{aligned} \qquad (5.30)$$

The left side corresponds to the differential voltage on the output of the detector from to the target and background signals. The right side is the differential voltage on the output of the detector from an equivalent blackbody signal that is close. The background temperatures are assumed to be identical in many cases, so the unknown quantity is the apparent temperature. Because (5.30) is transcendental, the solution is found using a table, trial and error, or some other numerical technique (i.e., there is no closed-form solution for exact apparent temperature).

5.3 Electro-Optical Targets and Backgrounds

The majority of target radiation whose wavelength is less than 3 to 4 μm is attributed to external radiation, or radiation from some other source. In the natural world of terrestrial objects, this radiation is reflected by targets and backgrounds.

The external radiation is the flux provided by the sun, moon, stars, skylight, and artificial sources. We assume that most objects are relatively opaque and diffuse. Under these assumptions, the primary target light flux that is viewed by an EO sensor comes from target reflections. The exception to these conditions is where the objects themselves provide visible light either by extremely hot blackbody emissions or by artificial means. Examples are rocket exhausts, fluorescent lamps, and laser emissions. With EO systems, the target-to-background signal is described primarily as a contrast. We first provide some background in sources and reflectivities and then describe the method for calculating contrast.

External Sources

The primary source for EO systems is the sun, which can be modeled outside Earth's atmosphere by a large 5,900K blackbody source. Figure 5.13 shows the irradiance of the sun both inside and outside Earth's atmosphere. Note the modification by the atmospheric transmission (given by the sea-level irradiance). The graph shown corresponds to the sun located directly overhead. For the sun located at other angles, the flux traverses a longer atmospheric path, reducing the irradiance even further. The sun's irradiance just outside Earth's atmosphere, for the mean Earth–sun distance, over all wavelengths is 1,390 W/m^2. Also, measurements of illuminance at sea level with the sun at its zenith (directly overhead) gives 1.24×10^5 lux. Illuminance for solar angles from the Earth's surface of 90, 45, 10, and 0 deg (sunset or sunrise) correspond to 12.4×10^4, 7.59×10^4, 1.09×10^4, and 732 lux, respectively.

An air mass m, corresponding to the secant of the angle from the zenith, can give a rough approximation of a corresponding irradiance reduction. This factor is the reduction factor for the solar irradiance across all wavelengths. More detailed atmospheric transmission is addressed in Chapter 6. Skylight, the light scattering from the atmosphere, also contributes to the target-and-background irradiance. Skylight becomes significant for objects in the shade and for large air masses. Skylight accounts for around one-fifth of the total illuminance at Earth's surface on a clear day. The treatment of skylight and cloud coverage can be found in Waldman and Wootton [10].

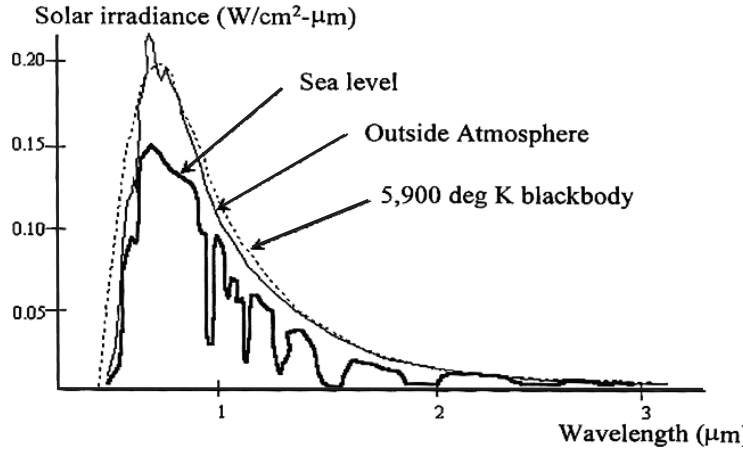


Figure 5.13 Solar irradiance at the Earth's surface.

The irradiance and illuminance provided by moonlight, or rather sunlight reflected by the moon, are related to a number of factors. The angular distance between the moon and the sun, the variation in Earth-to-moon distance (26% variation), the differences in the reflectance of the moon's surface (20% variation), and the angle of the moon above Earth's horizon all contribute to variations in moonlight. Moonlight illuminance ranges from 3.12×10^{-5} lux at worst conditions to 0.267 lux at its brightest.

Star brightness [11] is specified with a stellar magnitude that represents the visual brightness of a star at a point just outside Earth's atmosphere. A zero magnitude star corresponds to an illuminance of 2.65×10^{-6} lux. The true illuminance of a star referenced (divided) by the zero-magnitude illuminance (a nonzero quantity) is given by 2.512^{-sm} where *sm* is the stellar magnitude. Stars contribute 2.2×10^{-4} lux to the clear night sky. The illuminance provided by stars accounts for one-fourth of the light in the night sky with no moon. The rest of the light is provided by emissions from various atoms and molecules in the atmosphere. The spectral irradiance provided by the stars can be approximated with blackbody sources outside Earth's atmosphere. The *Electro-Optics Handbook* provides peak wavelengths for the generation of blackbody sources corresponding to various stars.

Human-made illuminators include tungsten lamps (2,700 to 3,300K), phosphors, sodium lamps, and mercury lamps. While some of these sources can be modeled with blackbody curves, they all produce a significant amount of light at EO wavelengths. The spectral emission of these illuminators can be found in a number of the handbooks and described in vendor references. The military spectral bandpass of choice for military EO (not infrared) systems is 0.65 to 0.90 μm. The human eye has little response within this band region.

For the typical solar, skylight, moonlight, starlight, and artificial illumination, reflectivity is the primary target parameter. The human eye is, in some ways, very similar to an EO sensor in that the variations in reflectivity drive the eye response. Different colors seen by humans correspond to the spectral reflectivity characteristics of objects. We use these characteristics to help us recognize objects. The exitance of light from an object is equal to the irradiance impinging on the object from some source, multiplied by the target reflectivity as a function of wavelength. An object that appears red has an impinging irradiance with red (around 0.6 μm) light content and the object reflectivity is higher at the red wavelength than at other wavelengths. Keep in mind that reflectivity can peak at a number of different wavelengths, where the wavelengths do not have to be within our visible response band of 0.4 to 0.7 μm. Therefore, objects can have peak reflectivities at wavelengths, or "colors," that humans cannot sense. Figure 5.14 shows the reflectivities for a number of backgrounds.

Three important spectral quantities affect the radiation leaving an object. The first is the source spectral irradiance $E_v(\lambda)$ of the object given a vacuum path (i.e., no atmospheric effects). The second is the atmospheric transmission between the source and the object $\tau_1(\lambda)$, and the third is the reflectivity of the object $\rho(\lambda)$. The exitance of the object is then

$$M(\lambda) = E_v(\lambda)\,\tau_1(\lambda)\rho(\lambda) \quad [\text{W/cm}^2\text{-}\mu\text{m}] \tag{5.31}$$

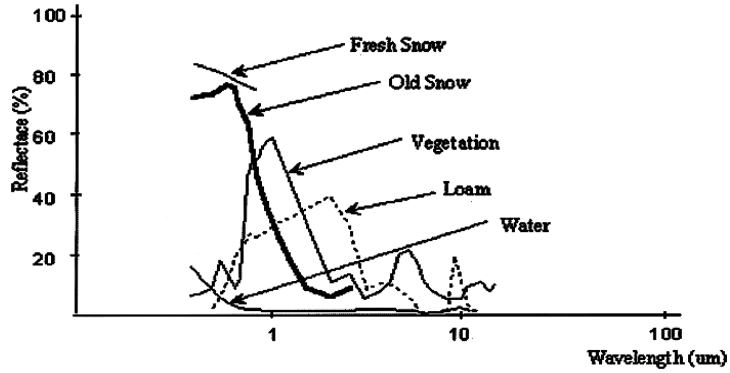


Figure 5.14 Reflectivity as a function of wavelength [10].

This exitance is a function of space over the object, especially because the reflectivity usually varies dramatically across object space. To determine the voltage from a detector within a given optical system from the object exitance described in (5.31), a few other parameters are required (Figure 5.15). These are the atmospheric transmission between the object and the sensor, $\tau_2(\lambda)$, the sensor transmission $\tau_{optics}(\lambda)$, and the detector responsivity $R(\lambda)$, in volts per watt. If the target surface is Lambertian, the radiance is calculated as $M(\lambda)/\pi$ in W/cm^2-sr-μm. The radiance can be multiplied by the source area seen by the detector abR^2/f^2, to obtain an intensity in W/sr-μm. The terms a and b represent the detector dimensions, f is the lens focal length, and R is the range from the target to the EO sensor. Finally, the intensity can be multiplied by the sensor solid angle $\Omega = \pi D^2/4R^2$ to obtain the flux on the detector. This combination of parameters over a given spectral bandwidth $[\lambda_1,\lambda_2]$ gives a detector voltage:

$$V = \frac{abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2} E_v(\lambda)\tau_1(\lambda)\rho(\lambda)\tau_2(\lambda)\tau_{optics}(\lambda)R(\lambda)\,d\lambda \quad [\text{V}] \tag{5.32}$$

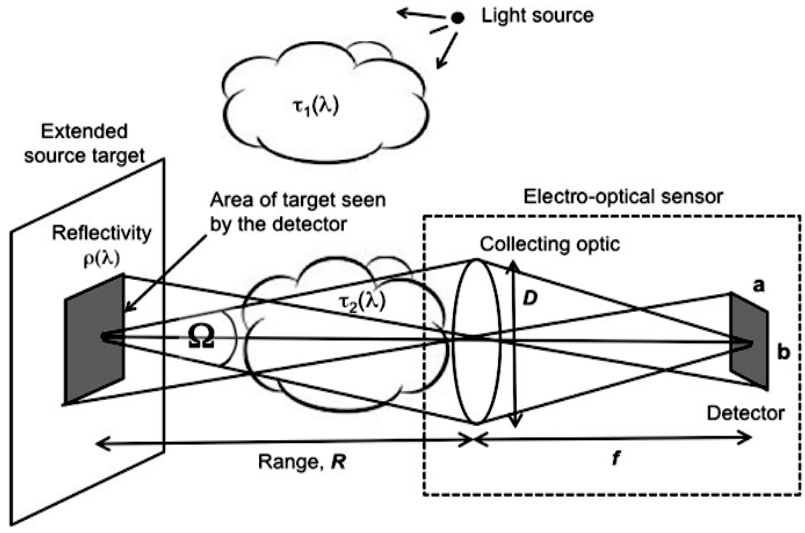


Figure 5.15 EO system light path.

This equation illustrates the complexity of EO systems. Unlike infrared systems, where the target is the source, a dual-path calculation is required. The dual paths are the source-to-target path and the target-to-sensor path.

All spectral modifications must be included to determine the correct voltage. An interesting phenomenon occurs when either (1) a target and background have completely different reflectivities (as a function of wavelength), but a common source spectrum, or (2) detector responsivity varies in a manner that causes identical voltage responses for the target and background. The multiplication and integration of (5.32) are identical for both the target and the background, but the reflectivities are different. In this case, the sensor cannot discern the target from the background. This phenomenon occurs occasionally with EO systems, but rarely with infrared systems.

To obtain the exact response of a system, all source, transmission, reflectance, and sensitivity quantities should be developed as a function of wavelength. The final response is then taken as the collective response (integrated) across all wavelengths. However, source irradiance, atmospheric transmission, reflection, and detector sensitivity are frequently specified as a band-integrated quantities [12]. For example, a silicon detector may have a sensitivity of 1,654 μA/lux between the wavelengths of 0.55 and 0.70 μm. While this quantity may not make sense, it is the ratio of the band-integrated source irradiance weighted by the spectral response of the silicon detector to the band-integrated source irradiance weighted by the response (efficacy) of the human eye:

$$R_{band} = \frac{\int_{0.55}^{0.70} S(\lambda)E(\lambda)\,d\lambda}{\int_{0.38}^{0.78} K(\lambda)E(\lambda)\,d\lambda} \quad \text{[A/lux]} \tag{5.33}$$

Note that the integration bands are not identical, causing some discrepancy in the notion of a reference. This type of band averaging and illuminance referencing is used widely and contributes to sensor response errors and limited accuracy. The errors can be reduced with a reduction in spectral bandwidth and a shift of the detector response to the visual response band. A high-fidelity analysis requires analytical terms to be band integrated after all responses have been included. For a quick-and-dirty analysis, band quantities are useful with reasonable results if the source irradiance, reflectivities, and atmospheric transmission do not vary significantly over the band.

Contrast

While all sensor systems respond to changes in radiation, EO systems have been modeled as responding to target-to-background contrast. This input parameter must be determined using the characteristics of illumination, target reflectivity, and background reflectivity. Contrast [13] is defined as the target-to-background luminance divided by the background luminance

$$C = \frac{L_T - L_B}{L_B} \tag{5.34}$$

The luminance quantities are in photometric units and are proportional to the detector output voltages. The weighting function that corresponds to the appropriate

light level of the human eye (i.e., photopic response), however, can be modified to the spectral response of the EO sensor. L_T becomes

$$L_T = \frac{abD^2}{4f^2}\int_{\lambda_1}^{\lambda_2} E(\lambda)\rho_T(\lambda)\,\tau_{atm}(\lambda)R_S(\lambda)\,d\lambda \tag{5.35}$$

where $E(\lambda)$ is the source irradiance on the target, $\rho_T(\lambda)$ is the reflectivity of the target surface, $\tau_{atm}(\lambda)$ is the atmospheric transmission between the target and the sensor, and $R_S(\lambda)$ is the spectral response of the sensor. The background luminance is calculated in the same manner.

The contrast perceived by the sensor is reduced from the value in (5.34) by *path radiance*, or stray light caused by the source illumination of particles in the atmospheric path. This modification results in the perceived contrast being

$$C = \frac{L_T - L_B}{L_B + L_{atm}} \tag{5.36}$$

where L_{atm} is the path radiance in the band of the sensor. This value is typically calculated with LOWTRAN or MODTRAN and can significantly affect the contrast.

Similar to the analysis of infrared systems, there are various levels of accuracy in analytical techniques. It is rare for (5.35) and (5.36) to be calculated in practice. More gross analysis is usually performed with a target-to-background contrast calculated at the target position as

$$C = \frac{\int_{\lambda_1}^{\lambda_2} E(\lambda)\rho_T(\lambda)R_S(\lambda)\,d\lambda - \int_{\lambda_1}^{\lambda_2} E(\lambda)\rho_B(\lambda)(\lambda)R_S(\lambda)\,d\lambda}{\int_{\lambda_1}^{\lambda_2} E(\lambda)\rho_B(\lambda)R_S(\lambda)\,d\lambda} \tag{5.37}$$

The source irradiance on the target has been measured for the sun, moon, and stars at Earth's surface for different seasons, climates, and times. Note the number of variables that are present in the analysis. These measurements allow consideration of the illumination geometry and atmospheric path between the illumination source and the target. The missing link between the contrast at the target position and the perceived contrast at the sensor is the path radiance. The contrast at the target position must be degraded by the atmosphere to obtain the perceived, or apparent, contrast.

Contrast at the target location can be calculated using various models. The U.S. Army Night Vision and Electronic Sensors Directorate (NVESD) has a contrast model that includes libraries of source irradiance, target and background reflectivities, and sensor-responsivity curves. These libraries show that a target and background can have dramatically different reflectivities but still produce a small contrast from the spectral characteristics of the source and sensor response.

Holst [13] approximates the perceived target-to-background contrast:

$$C_R = C\tau_{atm} \tag{5.38}$$

where τ_{atm} is the band-averaged atmospheric transmission approximated by

$$\tau_{atm} = e^{-\sigma_{atm}R} \tag{5.39}$$

Table 5.4 Scattering Coefficients

Condition	*Scattering Coefficient [km^{-1}]*
Dense fog	78.20
Moderate fog	7.82–19.6
Thin fog	1.96–3.92
Haze	0.98–1.96
Clear	0.19–0.39
Very clear	0.078–0.19

From Holst [13].

where R is the range in kilometers, and σ_{atm} is the atmospheric scattering coefficient. The atmospheric scattering coefficient is given for a number of conditions in Table 5.4.

Example 5.6

An armored personnel carrier (APC) travels along an Alabama clay road during a sunny day. Determine the target-to-background contrast of the vehicle using the human eye as the sensor. The normalized solar irradiance, the eye's normalized photopic response, the reflectance of the APC, and the reflectance of Alabama clay are given in Figure 5.16.

The eye's luminous response to the target is determined by the product of the solar irradiance, the APC reflectance, and the photopic response. The total luminous response is determined by integrating this product. The integral for the curves shown is 23.8 (normalized luminance). The eye's response to the background is determined by the integral of the product of the solar irradiance, the clay reflectance, and the photopic response. The normalized background luminance is 7.65. The contrast is determined using (5.37) to be 2.11 or 211%. Normalized values for the solar irradiance and the photopic response are acceptable as long as the target and the background are both evaluated as normalized quantities. Another widely used contrast is $C = (L_T - L_B)/(L_T + L_B)$, and in this example, results in $C = 0.51$.

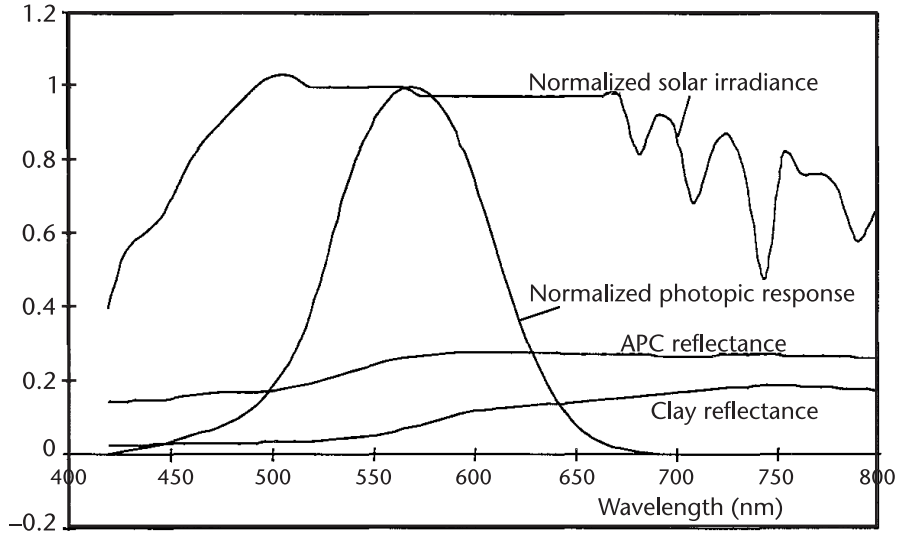


Figure 5.16 Contrast example.

5.4 Other Sensitivity Considerations

We have shown methods for calculating the power on detectors in the infrared and EO cases. The methods for determining a delta T for the infrared and a contrast for EO systems have been presented. Be careful, however, when using these quantities, because they are first-order approximations of target-to-background signals. Many assumptions accompany these calculations, and care should be taken not to apply these signal descriptions when the assumptions are not satisfied.

Bidirectional Reflectance Distribution Function

Although we have provided generalizations in order to obtain first-order characterizations of targets and backgrounds, an exact model of targets and backgrounds must address the reflectivity, emissivity, and transmissivity of every surface in the scenario. This is rarely performed by infrared and EO sensor developers; however, it is a common practice for target coating researchers. The Coatings Engineering Evaluation Program (CREEP) is an example of a model that was developed for the purposes of evaluating paints and coatings.

Programs such as CREEP derive directional parameters for infrared (and other band) emissivity and reflectivity [14]. The directional emissivity is given by

$$\varepsilon(\lambda) = \frac{L(\theta,\phi)}{L_{BB}} \tag{5.40}$$

where θ and ϕ define emissivity direction and L_{BB} is the radiance of a blackbody at the same temperature as the coating. The bidirectional reflectance distribution function (BRDF) is defined as

$$\rho_{BRDF} = \frac{L_r(\theta_r,\phi_r)}{E_i(\theta_i,\phi_i)} \tag{5.41}$$

where $L_r(\theta_r, \phi_r)$ is the reflected radiance and $E_i(\theta_i, \phi_i)$ is the incident irradiance. BRDF nomenclature is illustrated in Figure 5.17. Note that with reflectivity, there are both incident angles and reflected angles. The units of ρ_{BRDF} are reciprocal steradians. For a Lambertian surface,

$$\rho_{BRDF} = \frac{\rho_o}{\pi} \tag{5.42}$$

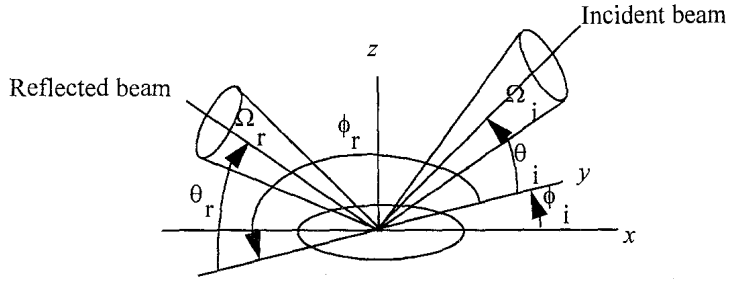


Figure 5.17 BRDF nomenclature.

where ρ_o is the reflectivity. If the surface is specular, the reflectance is only nonzero when $\theta_r = \theta_i$:

$$\rho_{BRDF} = R(\theta_i)\frac{\delta(\theta_r - \theta_i)\delta(\phi_r - \phi_i + \pi)}{cos\theta_r \, sin\theta_{kr}} \tag{5.43}$$

where $R(\theta_i)$ is the Fresnel reflectance. Recall that $\varepsilon = 1 - \rho_{BRDF}$ for a surface that does not transmit light. There are many models of BRDF for surfaces that are not Lambertian or specular. They can be thought of as surfaces between Lambertian and specular. Scientists attempt to optimize these surface coatings for particular applications. These applications include low observable (LO) technology, laser damage susceptibility, and radiation management.

Color Considerations

Many commercial EO sensors have separate detectors that correspond to red, green, and blue wavelength bands to obtain color information. These three bands have been shown, when used together, to give appropriate representation of the human visual response. Once detected, the three bands are combined and displayed using the color representation of images. Humans perceive the three colors to represent the entire visual spectrum. For most military systems, all of the light within a given band (not necessarily in the visual band) is integrated on a single detector response to give gray-level images. Highly reflective objects within a given band are represented by white (i.e., higher gray levels), where low-reflective objects are represented with dark gray levels. A single picture element (pixel) is typically represented by anywhere from 16 (4 bits) to 1,024 (10 bits) shades of gray.

Some new approaches in the use of reflectivity as a function of wavelength to detect, recognize, and identify targets are currently under investigation. For sensors that can detect flux as a function of wavelength (imaging spectrometers), neural networks are being used to discriminate targets based on their spectral signatures. In another approach, EO, midwave infrared, and longwave infrared bands are pseudo-colored with red, green, and blue colors for display to humans. The human training and human responses to these pseudo-colored images are yielding high probability levels of recognition and identification. One of the associated problems with this approach is the pixel registering of the three images (EO, midwave infrared, and longwave infrared). That is, each picture element must be aligned in the three images before the color coding and subsequent image display. As imaging systems become more sophisticated, spectral capabilities will be used frequently in tactical sensors. Currently there are few platforms that include extensive spectral segregation. Usually, a single-band EO sensor for day viewing and a single imaging infrared (midwave or longwave) sensor band – for night viewing are fielded on a weapons platform.

5.5 Target and Background Spatial Characteristics

We have previously discussed the radiometric characteristics of targets and backgrounds. In the infrared, the important target and background parameters are object temperatures and emissivities, whereas at EO wavelengths, they are the reflectivities

and external source irradiances. While the resultant object radiometry quantity is exitance as a function of wavelength, the exitance is also a function of space. In fact, an exact analysis of a sensor's response to a particular target and background requires an exact expression for the target and background exitance as a function of space, $M_\lambda(x,y)$.

Spatial exitance describes the plane of the target, but exitance can be represented in many ways. Three methods are common. The first is the exitance as a function of space in the plane of the target (normal to the sensor line of sight), where x and y are the horizontal and vertical independent variables, in meters. The second is in angular space where x and y are horizontal and vertical angular variables, in milliradians, that describe the azimuth and elevation points in the target plane as seen by the sensor. These angular variables are found simply by the horizontal and vertical positions in the target plane divided by the range from the sensor to the target. This angular method for describing targets is very useful in the linear systems approach to sensor analysis since spatial frequencies are typically specified in cycles per milliradian. Finally, another useful method of representing a target image is to write the irradiance pattern provided by the target and background on the sensor detector plane. The scaling of the $M_\lambda(x,y)$ function from the object plane to the image plane is performed simply by reducing the object size by the ratio of the effective sensor focal length to the target range $M_\lambda(xR/f,yR/f)$. Note that this scaling typically reduces the function by orders of magnitude in size and that the object exitance function must be multiplied by the appropriate radiometry parameters to convert the scene image to an irradiance in the detector plane.

The spatial frequency content, or spectrum, of a scene (target and background) can be found by taking the two-dimensional Fourier transform of the scene exitance

$$M_\lambda(\xi,\eta) = \mathcal{FF}\{M_\lambda(x,y)\} \tag{5.44}$$

where ξ and η are in cycles per meter. Note that the wavelength dependence is shown. The spectrum must be integrated over wavelength from the target exitance, through the atmosphere, through the sensors' optics, and onto the detector. With this in mind, the wavelength dependence is dropped and assumed to be given for some wave band of interest.

In a typical scene, a large number of frequency components are present. In practice, a fast Fourier transform (FFT) can be applied to the image to determine the spectrum of the image. Small objects have high-frequency components and large objects have lower frequency components. A common exercise for EO students involves a tank hiding within the leaves of a tree. An optical system is designed such that components of the tree leaves do not pass the optical system, but the significant tank frequency components are allowed to pass through the system. This reduction in resolution allows the viewer to quickly find the tank.

Angular space is used when an object position is specified in azimuth radians x and elevation radians y within an optical system's FOV. Spatial frequencies of the scene are then given in terms of cycles per milliradian. These are the typical quantities that are used in sensor and target analysis. A 3-m × 4-m truck located 1 km from a sensor would subtend 3 mrad × 4 mrad in angular space. Given that the

truck is modeled as a two-dimensional rectangle function $M(x,y) = rect\left(\frac{x}{3},\frac{y}{4}\right)$, the frequency components associated with the target are

$$M(\xi,\eta) = \mathcal{FF}\left\{rect\left(\frac{x}{3},\frac{y}{4}\right)\right\} = 12\ sinc(3\xi,4\eta) \tag{5.45}$$

where ξ and η are in cycles per milliradian. Oversimplification can occur by gross modeling of a truck with a large rectangle function. That is, the details of wheels, mirrors, headlights, doors, and so forth may not be resolvable given gross modeling. Sums of many different rectangles or other functions can give reasonable estimates. FFTs of real images taken with high-resolution imagers can give sufficient frequency component information to model targets and backgrounds. Other modeling includes facets of various shapes and sizes to construct a high-resolution target and background. The target spatial frequency spectrum is then the sum of the individual facet component spectra.

Bar Target Representation of Targets

An exact analysis of whether a human can recognize a given object with a sensor requires a spatial frequency analysis of the object and the sensor system. However, a technique that has been shown to provide reasonable results is that of *bar target representation* of targets by John Johnson [15]. It is common practice to represent objects with bar targets for the purpose of sensor evaluation, analysis, and design. Johnson conducted a number of experiments at NVESD to develop requirements for detection, recognition, and identification of objects (in particular, tactical ground targets). Consider the targets in Figure 5.18 along with their bar target equivalents. Johnson determined the number of bar pairs, or cycles, necessary across an object to perform a discrimination task. Further NVESD experimentation yielded similar results. The requirements for detection, recognition, and identification of targets are described in Table 5.5.

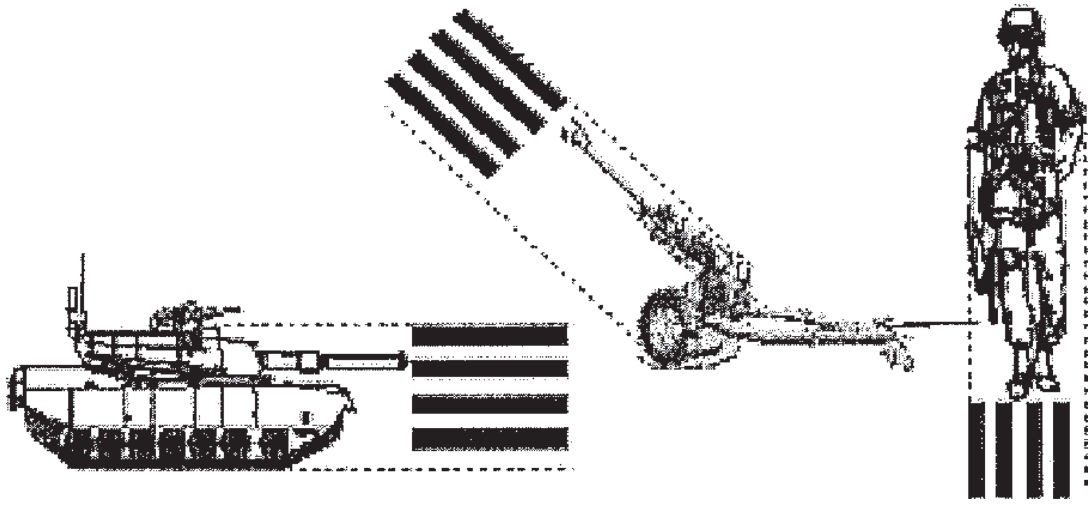


Figure 5.18 Objects and corresponding bar targets.

Table 5.5 Cycle Criteria

Task	*Description*	*Cycles Across Object Minimum Dimension* N_{50}	*Cycles Across Two-Dimensional Object* N_{50}
Detection	Reasonable probability that blob is the object being sought	1.0	0.75
Recognition	Class discrimination (human, truck, tank, and so on)	4.0	3.0
Identification	Object discrimination (M1A, T-62, or T-72 tank)	6.4	6.0

The one-dimensional description corresponds to the number of cycles across the minimum dimension of the object. A two-dimensional cycle requirement is given for the characteristic target dimension (a function of target width and height as seen by the sensor):

$$h_c = \sqrt{W_{tgt} H_{tgt}} \quad [\text{m}] \tag{5.46}$$

These criteria are valid for tactical targets that have a width-to-height aspect ratio of less than 3 or 4. They do not apply to long, thin targets. The cycle criteria in Table 5.5 correspond to a probability of 50% for the discrimination task given (thus, the N_{50} nomenclature). If other probabilities are desired, Holst [16] describes the TTPF as a multiple of the number of cycles given in Table 5.5 to achieve a particular probability of discrimination. The TTPF is given in Table 5.6. For example, recognition requires three cycles across a two-dimensional target to assess the probability of recognition at 50%. To provide a recognition probability of 95%, three cycles times the 2.0 TTPF factor yields a requirement of six cycles across the target. The equation for the TTPF is provided later in this book in the systems section.

While these resolution tasks are necessary for the discrimination of targets, sufficient sensitivity is assumed. Do not forget that both sensitivity and resolution are the performance requirements necessary to address any sensor task (given that basics such as size, weight, power, and so forth have been met). Questions that should be discussed before any sensor design begins are as follows: What are the flux characteristics of all objects that could be in the scene of the target and of the background? What are the possible target sizes and shapes? What are the resulting resolution requirements (sensitivity and spatial resolution) necessary to discriminate

Table 5.6 Target Transfer Probability Function

Probability	*Multiplier*
1.00	3.00
0.95	2.00
0.80	1.50
0.50	1.00
0.30	0.75
0.10	0.50
0.02	0.25
0.00	0

these targets? The target and background issues must be defined to proceed with a sensor design and/or analysis for any given scenario.

Target Delta T and Characteristic Dimension

The target delta T (ΔT), sometimes called the RSS delta T for the root-sum-squared delta T, is used to describe the thermal contrast of a target. When this value is measured in the field (as shown in Figure 5.19), the target and background picture elements (pixels) are first calibrated with reference blackbodies. Each pixel that usually provides gray levels associated with the brightness of the infrared image are converted to an equivalent blackbody temperature as seen by the pixel. That is, each pixel value is given in blackbody temperature (Celsius or kelvin).

The target is segmented as shown in Figure 5.19. The background is taken as the region around the target and is sometimes specified as a box with the same number of pixels as that of the target area. Occasionally, it will be specified as an area twice the number of pixels as the target area. If the target area is converted from number of pixels to square meters, then the characteristic dimension of the target is the square root of the target area in meters. Note that the characteristic dimension changes with aspect angle of the target (front is small and side is large).

The target delta T is

$$\Delta T = \sqrt{(\mu_{tgt} - \mu_{bkgd})^2 + (\sigma_{tgt})^2} \quad \text{[K or °C]} \tag{5.47}$$

where μ_{tgt} is the mean of the target pixel values in degrees Celsius, μ_{bkgd} is the mean of the background pixel values in degrees Celsius, and σ_{tgt} is the variance of the target pixel values in degrees Celsius. Note that it does not matter whether you use kelvin or Celsius because the delta T values are differential.

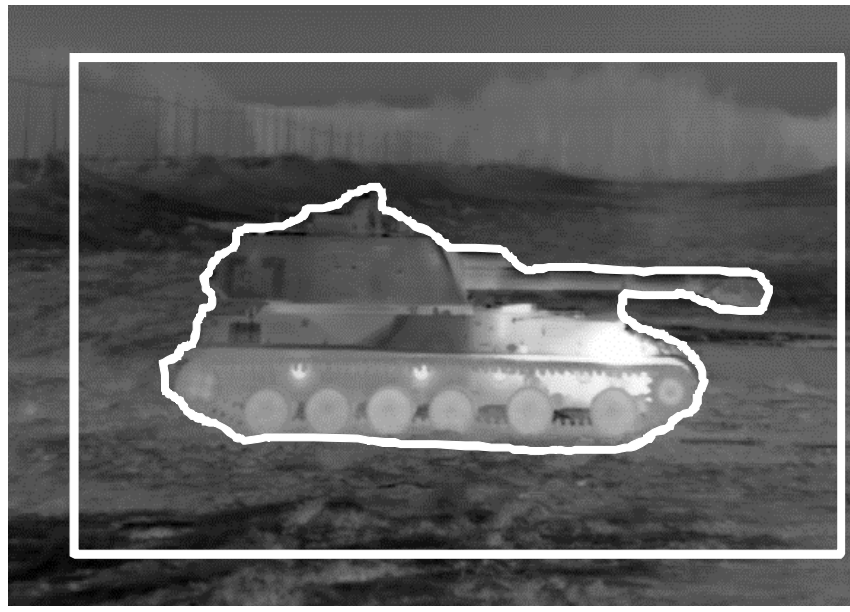


Figure 5.19 Ground target, segmentation, and background.

Table 5.7 Vehicle Target Characteristics

	Band	*Char Dim (m)*	RSSΔT *(C)*	*Contrast*	N_{50}	V_{50}
Military Vehicle Recognition	LWIR	3.65	3.4	0.18	4.4	16.4
Tracked Armored Vehicle Identification	Visible	3.0		0.28	7.5	22
	LWIR	3.11	3.4	0.205	8.0	20.5
Commercial/Paramilitary Vehicle	MWIR	2.13	7.0	0.35	7.0	28.0
Identification	LWIR	2.13	7.3	0.4	6.0	27.5
Wheeled Vehicle Identification	Visible	3.3		0.265	3.4	8.5
(Air to Ground)	Thermal	3.3		0.389	6.0	25.0
Wheeled Vehicle Identification	Visible	2.7		0.313	2.8	8.5
(Ground to Ground)	Thermal	2.7		0.415	7.7	32.0

Summary of Target Characteristics

Although the discrimination criteria provided in the previous section are frequently called the Johnson's criteria, they represent a historical quantity that is sometimes still used, but a similar quantity V_{50} is an integrated metric that describes the target discrimination difficulty of an integrated set of frequencies (integrated in contrast and spatial frequency). Three quantities describe targets in the target acquisition process: target contrast, characteristic dimension, and the discrimination criteria V_{50}. To be complete, we include N_{50}. Table 5.7 is a useful reference for vehicle

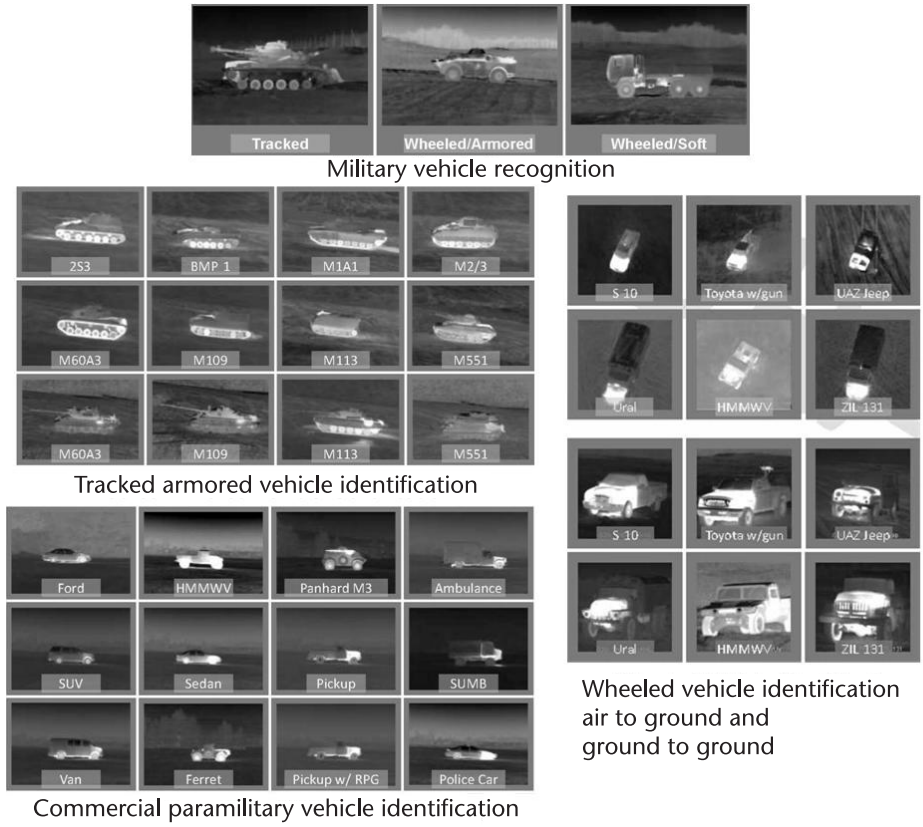


Figure 5.20 Vehicle target discrimination definitions (see Table 5.7).

target quantities in target acquisition calculations. We describe how these quantities are used in the later systems chapters. The band is given, the characteristic dimension is the root of the average area of the target, the RSS delta T in degrees Celsius is described in the next section, contrast is the visible contrast of the target, and the N_{50} and V_{50} values are discrimination difficulty parameters. The discrimination difficulty parameters are described in more detail in the systems chapters. The blank cells of the table mean that the parameters are not available at this time and may be available in the future.

The discrimination tasks provided in Table 5.7 are more fully described in Figure 5.20. The N_{50} and V_{50} values given are the discrimination values required for an average observer to obtain a 50% probability of discrimination level. That is, the observer would be able to discriminate between the vehicles shown in the definitions to a 50% probability of discrimination level. Figure 5.21 shows the static object discrimination definitions for Table 5.8. The dynamic definitions are not shown in a collage; however, the 12-target urban activity discrimination is a *dynamic* discrimination target characteristic where human activities can be determined. For the 12-target urban activities, examples are reading a book, talking on a cell phone, positioning a rocket-propelled grenade launcher, and 9 other dynamic activities. The 24 rural activities include sweeping, raking, holding a rifle, and 21

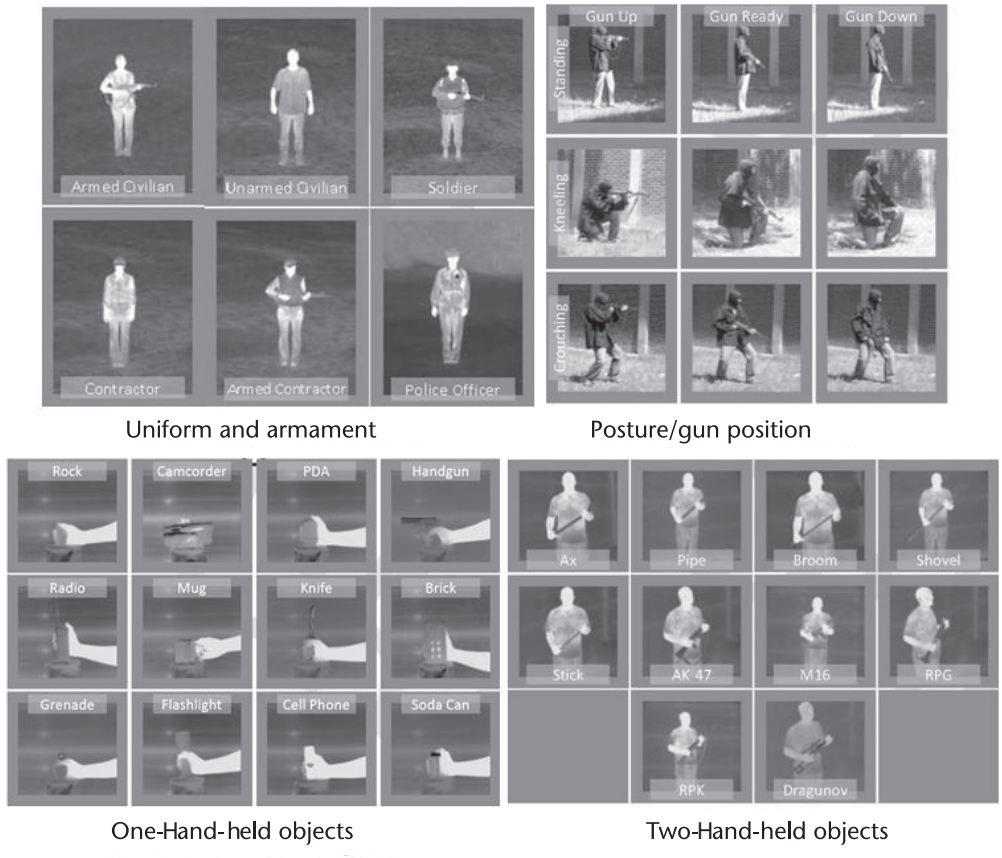


Figure 5.21 Other target discrimination definitions (see Table 5.8).

Table 5.8 Other Target Characteristics

	Band	*Char Dim (m)*	RSSΔT *(C)*	*Contrast*	N_{50}	V_{50}
Uniform and Armament	MWIR	0.7	5.4	0.4	3.25	15.0
	LWIR	0.7	5.0	0.45	3.7	18.8
Body Posture & Gun Position	Visible	0.99		.29	7.0	19.4
	MWIR (Day/Night)	1.06	4.2	0.15	6.5	15.0
		0.90	1.6	0.20	3.2	6.6
	LWIR (Day/Night)	1.09	5.2		4.9	11.4
		0.84	2.1		2.7	6.6
One-Hand-Held Object	Visible	0.1		0.3	3.5	9.5
	MWIR	0.1	3.3	0.25	4.7	18.0
	LWIR	0.1	3.0	0.26	4.6	17.0
Two-Hand-Held Object	Visible	0.25		0.2	2.7	9.5
	MWIR	0.25	4.1	0.24	4.3	16.0
Activity (12 target urban)	MWIR (static)	0.73	2.89	0.37	7.2	35.4
	MWIR (dynamic)	0.73	2.89	0.37	3.2	11.7
Activity (24 target rural)	Visible	0.70		0.38	2.6	9.6
	SWIR (day)	0.70		0.36	3.25	12.8
	MWIR (day/night)	0.70	5.34	0.60	2.6	12.2
		0.70	4.52	0.41	3.1	12.2
Maritime Activity	MWIR	0.62	1.97	0.24	1.6	7.6

other dynamic activities. Finally, the maritime activity discrimination includes fishing, dropping an anchor, and 10 other maritime activities.

The target discriminations tasks and the associated target characteristics are all used to model, calculate, and determine how far an average observer with an infrared or EO sensor can perform the discrimination at a given range (distance to the target). The characteristic target dimensions, target contrasts (for visible), and delta T's (for infrared) are measured directly from the target with on-site radiometers and are averaged over the target set and target viewing angles to give nominal values. The N_{50} and V_{50} are determined with psychophysical experiments to provide probabilities associated with different levels of information across the target. These experiments are performed to provide discrimination calibrations that can be used in target acquisition calculations. Finally, while small boat targets are not provided in the tables, the characteristic dimension is 3.9m, the contrast is 0.3 in both the visible and SWIR, and the delta T is 7.3K in the day and 1.6K at night in the MWIR. The N_{50} is 4.0 for the visible, 3.5 for the SWIR, 2.8 for the MWIR day, and 4.0 for the MWIR night. The V_{50} is 14 for the visible, 11.3 for the SWIR, 10.6 for the MWIR day, and 13.6 for the MWIR night.

Clutter

As discussed in the introduction, clutter [17, 18] describes the background components that resemble the targets in size and shape. Clutter has been modeled by lowering the effective contrast of the target and the background and by edge contrast modification in a scene. A common model that is used frequently is one where the clutter is defined as

$$CL = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2} \tag{5.48}$$

where the scene is divided into N blocks. The blocks are selected to be twice the size of the target (in pixels). The variance of the pixel intensities within block i is given as σ_i^2. The clutter is taken as the square root of the normalized variance as shown in (5.48). Finally, the signal-to-clutter ratio (SCR) is defined as the target-to-background contrast divided by the clutter.

The effect of clutter on human target acquisition capability is frequently modeled by three categories: low, moderate, and high. The TTPF is modified for these clutter levels to decrease the acquisition probability. Instead of modifying the TTPF, Holst suggests that the changes in TTPF can be approximated simply by multiplying the cycle criteria of Table 5.6 by the values given in Table 5.9. Note that the previous acquisition model was developed for the case of moderate clutter (i.e., a multiplier of 1). Most natural scenes fall into the moderate-clutter region, however, some urban scenes correspond to the high-clutter region. Some desert and calm ocean scenes place clutter levels in the low region.

Simulation of Target Characteristics

Modeling and simulation have been—and still are—a major thrust since the late 1990s, particularly for the military industry, in which shrinking defense budgets require more modeling and simulation efforts than ever before. Live testing is being reduced and augmented with simulations. Not only are target engagements using particular sensors being modeled, but entire battles that include surveillance and fire control sensors are under development. Questions for the sensor designers include these: How well and how quickly can I detect and classify what I am seeking? Can I hide from what is seeking me? Additional questions that are asked by military leaders include these: Can this sensor significantly improve my chances of winning the battle? Can I use the technology to save human lives? Modeling and simulation, if performed correctly, can provide answers to many of these questions.

Target simulations must provide realistic conditions such that sensor system performance can be evaluated when engaging the target. The conditions include both geometric and exitance fidelity of the target. For infrared systems, the exitance characteristics include the target temperature and the wavelength dependency in the appropriate wavebands. In some cases, solar reflections must be addressed. EO systems require characterization of target reflectance and illumination parameters. The geometric fidelity requirement can be derived from the sensor resolution and the apparent temperature difference between resolution elements on the target.

A number of different methods are used to model targets, including modeling only the target exterior surface (boundary representation) or modeling entire target volumes (solids modeling). The temperature and emissivity modeling or reflection and illumination modeling can be implemented from actual target measurements or from first principle calculations. Advantages of first principle calculations are that one can apply greater flexibility to operating conditions, whereas the disadvantage is that they result in greater inaccuracies than measured data. Not only are models of hardbodies required, but rocket plumes, jet engine exhausts, and combustion engine exhausts are also needed.

Table 5.9 Clutter Modifications to N_{50}

Clutter	*SCR*	N_{50} *Multiplier*
Low	>10	0.50
Moderate	1 < SCR < 10	1.00
High	<1	2.50

A large number of programs are currently available for target modeling. They are more popular in the infrared systems because infrared testing is much more expensive than EO testing. A partial list of target models include the following:

AEMAT	AvLab Electro-Optical Model for Aerial Targeting
EMAT	Electro-Optical Model for Aerial Targeting
GTSIG	Georgia Tech Thermal Contrast Model
IRMA	Infrared Modeling and Analysis
PCNirATAM	NATO Infrared Air Target Model
PRISM	Physically Reasonable Infrared Signature Model
SIRIM	Simulated Infrared Image Model
SPIRITS	Spectral and In-Band Radiometric Imaging of Targets and Sensors Model

Many of these models are owned by the U.S. government and can be obtained with government agreement. See *Infrared Signature Simulation of Military Targets* [19] for further information.

5.6 Typical Midwave and Longwave Contrasts and Solar Effects

The RSSΔT has been used previously as a means of comparing the respective abilities of MWIR and LWIR sensors to distinguish various targets [20]. In the cited work, both a T-72 tank and a Humvee (HMMV) were selected as targets. Simultaneous MWIR and LWIR images of each target were collected every 2 hr starting approximately 1 hr before sunrise and ending approximately 2 hr after sunset. Prior to data collection, the target vehicles were exercised by driving for at least 20 min. Data were also acquired from four different blackbody sources for converting the observed photon counts to temperature.

Figure 5.22 shows sample imagery collected from both the HMMV (left) and T-72 tank (right) in both the midwave and longwave regions. A comparison of the MWIR and LWIR RSSΔT is shown in the scatterplot of Figure 5.23. Each point on the graph represents the simultaneous MWIR and LWIR signatures at a particular aspect angle and time of day. For a given measurement, if both RSSΔTs are the same, that point would fall on the diagonal indicating an equal ability to resolve the target. While time of day was recorded (and indicated in the plot), aspect angle was not and is one of the primary causes of the "scatter" of points on the plot for a given time of day.

For the HMMV target, the distribution of RSSΔT in Figure 5.23 appears to indicate that over the entire day there were more images in which the LWIR signature

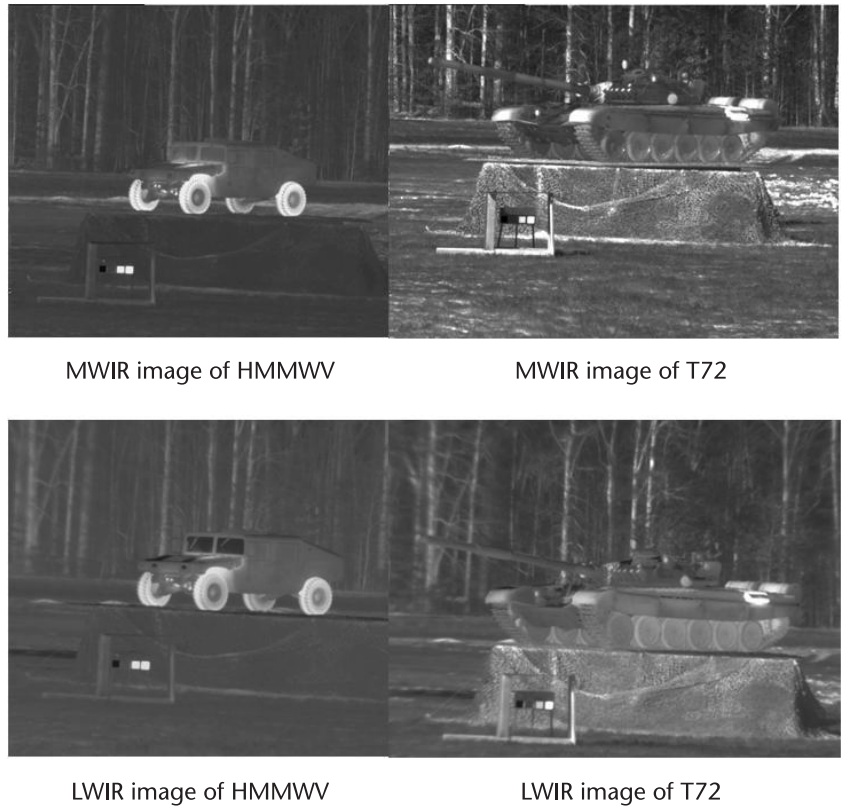


Figure 5.22 MWIR and LWIR images of a HMMV and a tank.

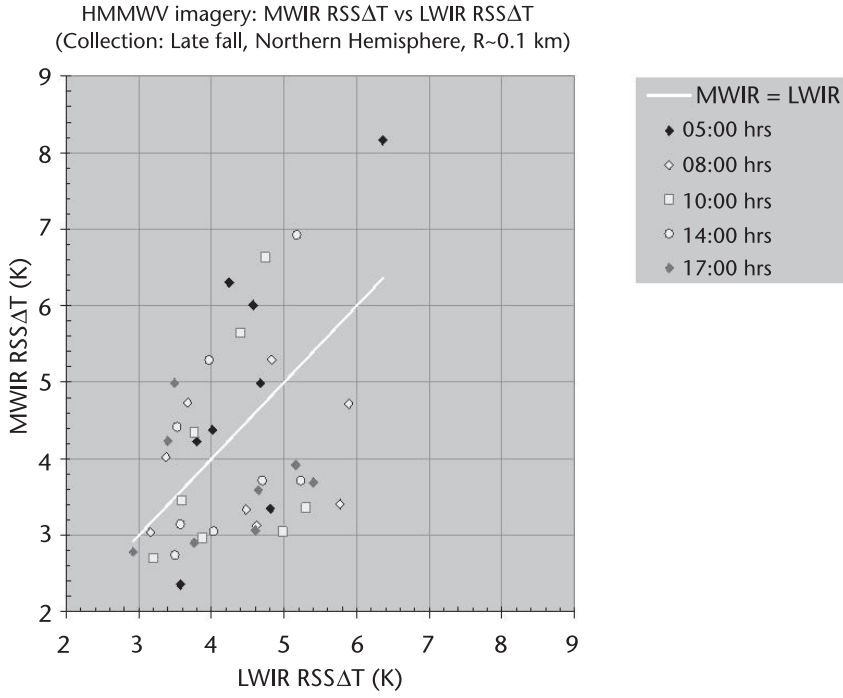


Figure 5.23 Scatterplot of HMMV MWIR RSSΔT vs. HMMV LWIR RSSΔT.

was larger than that of the MWIR. When averaged over the entire day, the LWIR signature exceeds the MWIR signature by ~4% of the average HMMV signature. Figure 5.23 also shows that before the sun came up the HMMV MWIR signature started out significantly larger than the LWIR signature. The fact that the HMMV signature was dominated by the MWIR during the earlier hours may be partly due to its windows, which are highly reflective in the IR. The MWIR sensor had no spectral notch filter to block the path radiance from the atmosphere (significant CO_2 absorption and emission between 4.2 and 4.2 μm), and thus the windows look warm in the MWIR image rather than cold as they do in the LWIR image.

For the T-72 target, the distribution of RSSΔT in Figure 5.24 shows that over the entire day there were more images in which the LWIR signatures were larger than the MWIR signatures. The LWIR RSSΔT averaged over the entire day exceeds that of the MWIR by 10% of the average T-72 signature.

The analysis of the RSSΔT signatures of the HMMV and T-72 conventional targets against a rural background indicates that when considered on an overall average basis, there was little difference between the MWIR and the LWIR signatures. However, when the signature data are compared on a time of day, vehicle type and size, and aspect angle basis, these results do not support the assessment that one band is superior for all situations. Figures 5.23 and 5.24 show that there is significant scatter between the MWIR and LWIR signatures, and it is this scatter with both bands present that greatly increases the likelihood of a large, significant target signature occurring in at least one of any pair of MWIR and LWIR images.

Also studied were the RSSΔT values for human targets against a rural background. All imagery was collected in July 2003 during a warm clear night. Each

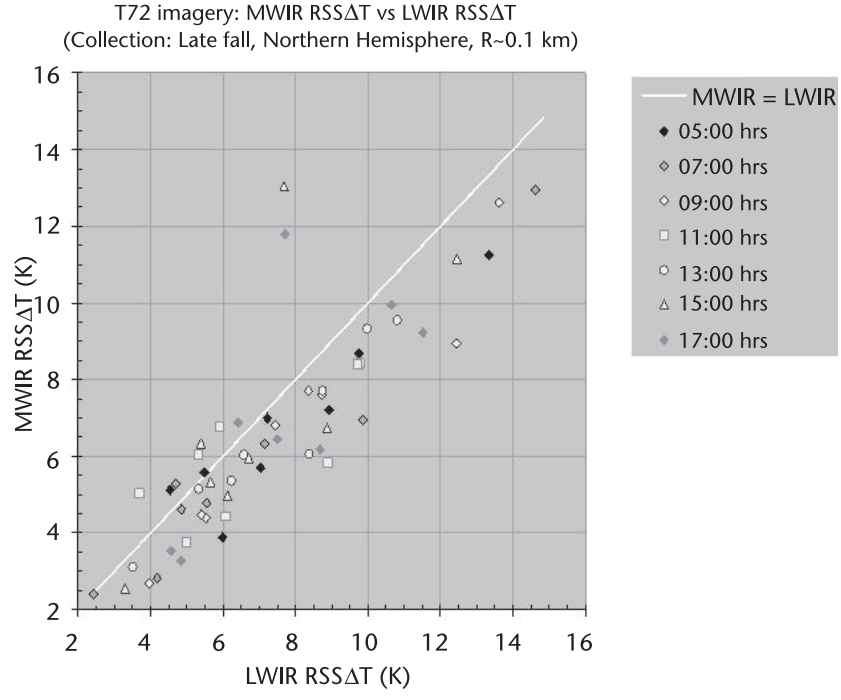


Figure 5.24 Scatterplot of T-72 MWIR RSSΔT vs. T-72 LWIR RSSΔT.

sensor had a slightly different FOV, but for both the target was positioned at a range of 12m so that the target filled a significant portion of the FOV in both cases. The short distance to the target also eliminated concerns regarding atmospheric extinction and path radiance.

A single human target modeled a set of uniforms and weapons: armed civilian, unarmed civilian, unarmed contractor, armed contractor, armed soldier, and armed police. As armed and unarmed civilian models, the subject was clothed in a lightweight cotton shirt with heavy cotton trousers. As armed civilian models, the subject carried an AK-47 rifle. As an unarmed contractor, the subject was clothed in cotton button-down shirt and lightweight cotton trousers while the "armed contractor" model added body armor and an M16 rifle. As a police model, the subject was clothed in a polyester uniform complete with badge and side arms. Finally, as a soldier model, the subject was clothed in a U.S. Army battle dress uniform (BDU) with rifle, Kevlar helmet, and body armor. A male, armed contractor is shown in both LWIR and MWIR in Figure 5.25.

A total of 144 different images were collected in each waveband (MW and LW) at a variety of angles in azimuth and cardinal headings. The human target and background regions of interest (ROIs) were manually segmented from each other with the background ROIs having the same number of pixels as the target ROIs. For each target, uniform, and aspect angle, target temperature variation, the RSSΔT was computed. As with the vehicle example, results are shown in scatterplot format in Figure 5.26. There appears to be a slight bias in favor of MWIR in the respective abilities of the two sensors to resolve human targets. For a given subject model (e.g., armed contractor) there was not much scatter between the MWIR and LWIR signatures.

In understanding target-to-background contrast it is also helpful to understand how a typical background varies spatially. The background spatial variance, σ_T, does not appear in the RSSΔT metric, yet can clearly impact one's ability to refind a target. To better understand the σ_T for various backgrounds, imagery was collected in both the MWIR and LWIR containing a mixture of rural and urban materials. Using the same two sensors as those used in the aforementioned human RSSΔT study, data collection was performed at three different times of day over



Figure 5.25 Image of male armed contractor: (left) LWIR and (right) MWIR.

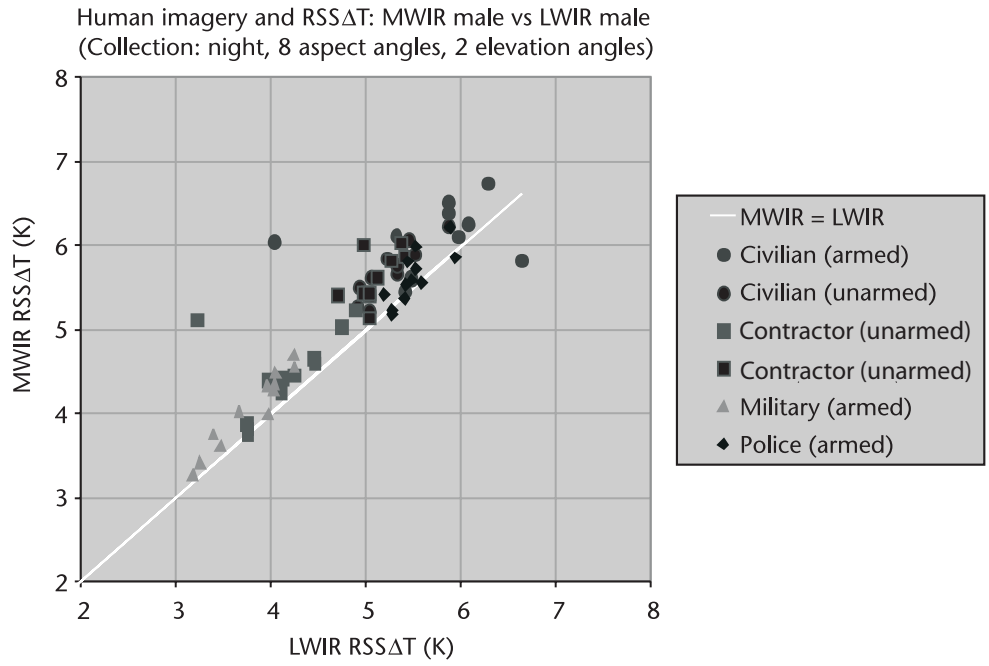


Figure 5.26 Scatterplot of human MWIR RSSΔT vs. LWIR RSSΔT.

an 8-month period (April–November 2004). The background materials considered were grass, trees, asphalt roads, asphalt shingles, white-painted concrete block walls, gravel, water, window glass, and a painted door. A given scene was always imaged from the same vantage point and the ranges varied from 100 m to 300 m, depending on the material being imaged. As before, ROIs were manually selected from each of the recorded images.

Figure 5.27 shows sample imagery from the various backgrounds in both the midwave and longwave regions. Scatterplots similar to those in the preceding examples are shown in Figure 5.28 for all data collected at 1400 hours. As before, points falling near the diagonal indicate an equal background σ_T for both MWIR and LWIR sensors. When considering all data, the rural background materials as seen through the LWIR sensor possessed a 0.06K greater σ_T than the MWIR sensor. For the urban backgrounds a 0.04K increase in σ_T over the MWIR sensor was observed. However, average σ_T is not necessarily the best statistic to focus on. From the observed scatter in Figure 5.28 we can see that certain materials, imaged at certain times of day can yield substantial differences in background pixel variance.

Figure 5.29 shows the same type of scatterplot for urban backgrounds. Note that the spread of σ_T is larger than that of the rural backgrounds because roads, roofs, and windows have more contrast than objects in a rural setting.

An additional issue that is frequently overlooked in background-to-target contrast is the influence of solar loading (heating) on both LWIR and MWIR signatures. When the sun is illuminating a scene, some regions are in the sun and some are in shadow. As previously shown and discussed, reflected sunlight can contribute significantly to MWIR images but almost nothing to LWIR images. The sun heats the nonshaded portions of the scene; subsequent increases in temperature in

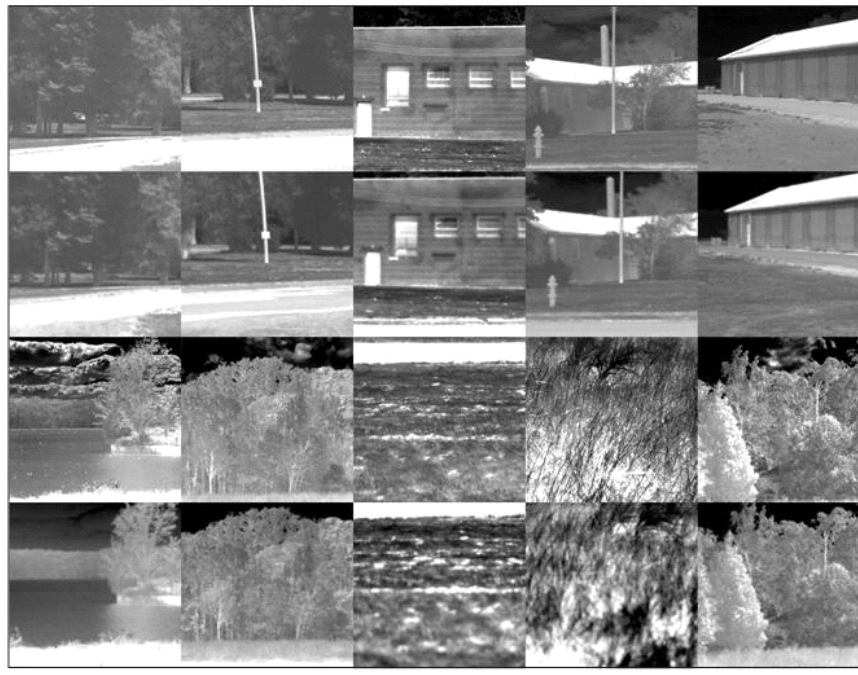


Figure 5.27 Collage example of MWIR and LWIR background images. (In each vertical pair of images of the same scene, the upper image is the MWIR image and the lower is the LWIR image).

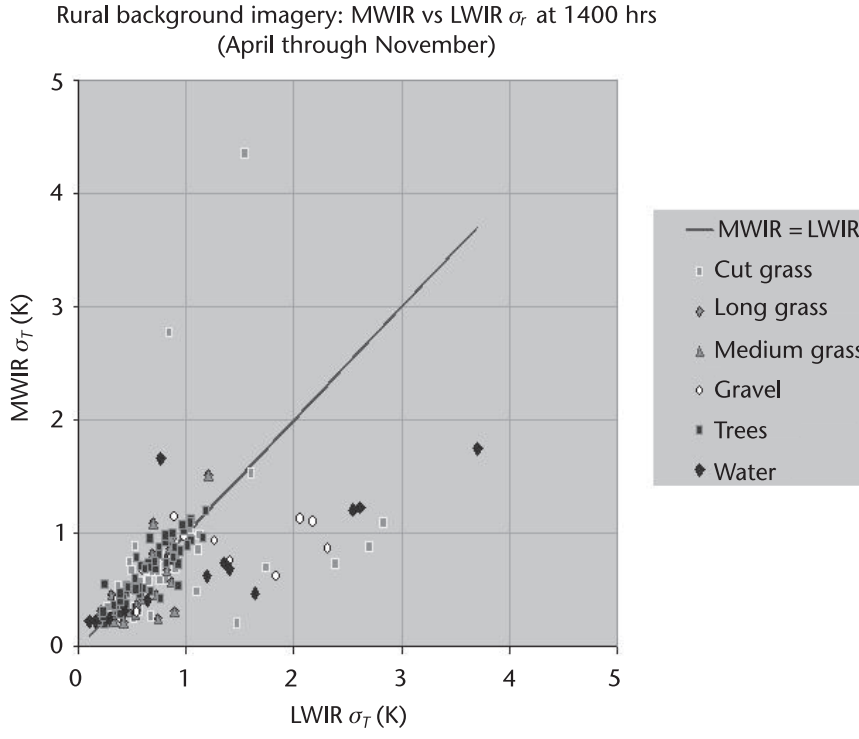


Figure 5.28 Scatterplot of rural MWIR σ_T versus LWIR σ_T.

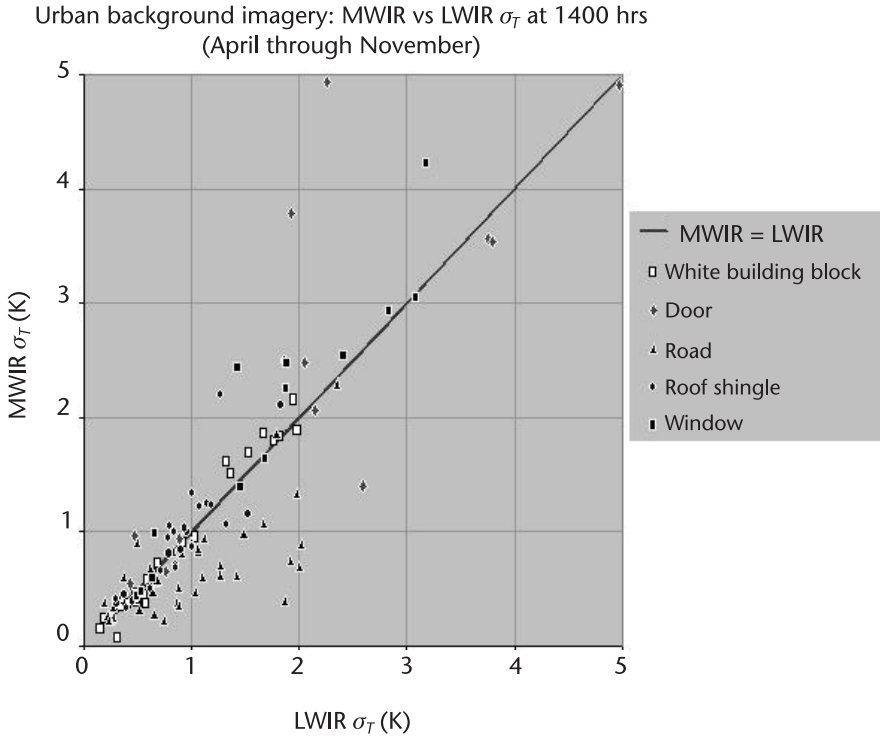


Figure 5.29 Scatterplot of urban MWIR σ_T versus LWIR σ_T.

these portions will contribute to both MWIR and LWIR signatures. To attempt to measure how much of the sun's effect is due to reflection and how much is due to loading, a simple experiment is provided.

A set of sandboxes was constructed. Each sandbox contained areas exposed to the sun as well as regions of permanent shade and temporary shade. The boxes were exposed to the sun for a couple of hours to allow the permanent shade region to equilibrate to the ambient air temperature and the no-shade region to reach

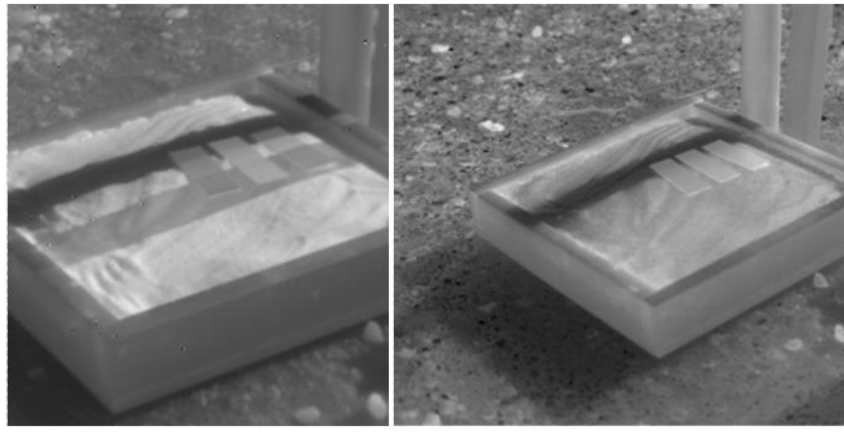


Figure 5.30 Sandbox images: (left) MWIR and (right) LWIR. The board on left is long term (solar loading) and the board on right is short term (solar reflection).

Table 5.10 Thermal and Solar Reflection Contributions

			Real Temperature (°C)		*Apparent ΔT (°C)*		
Material	*Method*	*Band*	*Air*	*SE Region*	*Scene*	*Sun Reflected*	*Solar Load*
Sand	Measured	MWIR	14*	25**	13.6	3.8	9.8
Concrete	Modeled	MWIR	14*	25	13.8	2.8	11.0
Sand	Measured	LWIR	14*	25**	9.8	0.4	9.4
Concrete	Modeled	LWIR	14	25	10.9	0.01	10.89

*Direct measurement; **Estimated

some stable heated state due to the sun. During this period of time, the boxes were almost continuously rotated manually to keep the permanent shade over the same section of sand. The boxes were then simultaneously imaged in the MWIR and LWIR with the same sensors used in the earlier examples of this section. An additional small portion of the sandbox was then temporarily shaded and another pair of MWIR and LWIR images was quickly acquired. The temporary shade was removed and the experiment was repeated after some period of time. Figure 5.30 shows a pair of MWIR and LWIR images of a sandbox with both the permanent and temporary shading structures in place.

The difference in apparent ΔT between the shade and the sun in the sandbox was both measured and modeled (using the temperature and emissivity of the sand). The solar reflection was also included in the modeling with the solar reflectivity in the MWIR and LWIR. Table 5.10 shows the results of the measurements and modeling. The modeled values were calculated for a concrete surface where the emissivity and reflectivity are very close to those of sand. The results show that the solar reflection contribution of an infrared signature can be 20% to 30% of the entire signature in the MWIR, whereas solar reflection contributions are negligible in the LWIR. This is one reason that automatic target recognizer (ATR) work favors the LWIR where there is little solar variation in the signature due to reflected light.

5.7 Exercises

5.1 A light source delivers 3W of power uniformly over an entire sphere. What is the source intensity?

5.2 The power of sunlight is measured in a 1-m2 area on Earth's surface and is 1,353W. Find the intensity of the sun if the distance from Earth to the sun is 150,000,000 km.

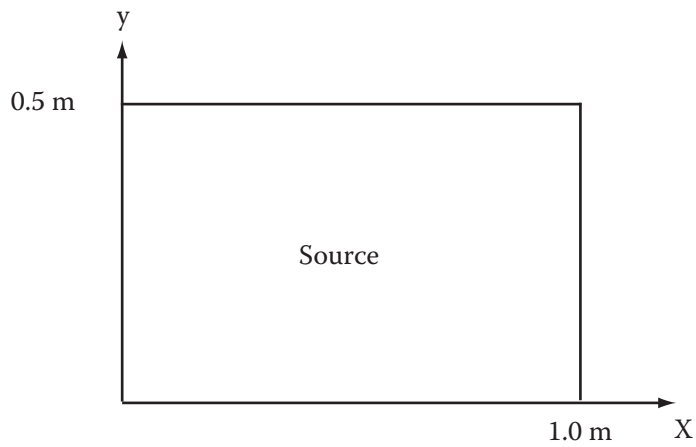
5.3 A laser delivers 5W of power in a beam with a 1-mrad divergence angle. Determine the laser intensity.

5.4 5.4 A spherical ball with a 1-cm radius emits 2W of radiation. Determine the average exitance of the sphere.

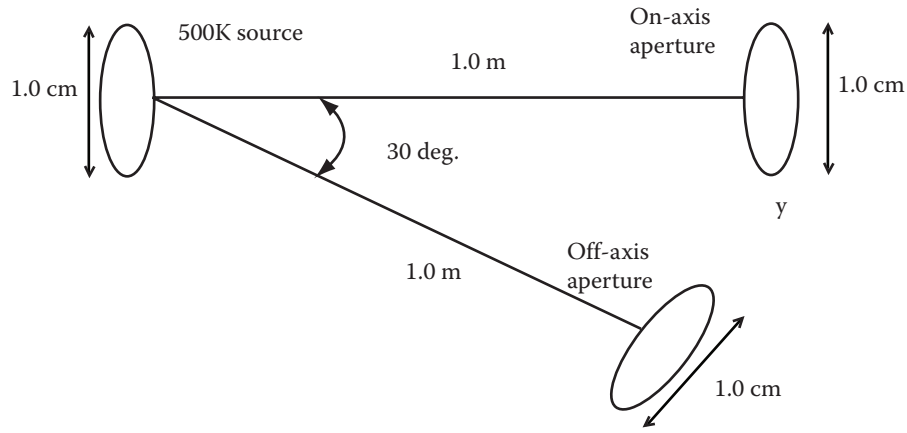
5.5 A neutral density filter is placed in front of a square light source as shown below such that the exitance of the source/filter combination is

$$M = 5x \quad [\mathrm{W/m^2}]$$

Determine the total power emitted by the combination.



5.6 A Lambertian blackbody source is heated to 500K as shown. An aperture is placed 1m away from the source, first on axis and then 30 deg off axis. Determine the radiance of the source, the power through the on-axis aperture, and the power through the off-axis aperture.



5.7 Determine the exitance and radiance of a 773K blackbody source:
a) over all wavelengths.
b) for the 3- to 5-ηm wavelength band.
c) for the 8- to 12-ηm wavelength band.
d) determine the peak emission wavelength.
e) Repeat (a) through (d) for a 300K source.

5.8 Calculate the difference in radiance for a 275K background and a 300K target. Repeat for a 300K background and a 325K target. Provide these calculations first for all wavelengths, then for the midwave and longwave bands. Does a 25K difference in temperature always give the same difference in radiance?

5.9 Provide the cycle criteria necessary to recognize a target with a 90% probability of recognition. Provided the target is 4 m × 6 m in size, what is the required angular spatial frequency that must be resolved by the sensor?

References

[1] *The Photonics Dictionary*, Pittsfield, MA: Lauren Publishing Company, 1996, p. D–97.

[2] Pinson, L., *Electro-Optics*, New York City, NY: Wiley, 1985, p.18.

[3] Jacobs, P., *Thermal Infrared Characterization of Ground Targets and Backgrounds*, Bellingham, WA: SPIE Press, 1996, p. 22.

[4] Wolf, W., and G. Zissis, *The Infrared Handbook,* Environmental Research Institute of Michigan, Office of Naval Research, Washington, DC, 1993, pp. 1–18.

[5] Pinson, L., *Electro-Optics*, New York City, NY: Wiley, 1985, p. 15.

[6] Waldman, G., and J. Wootton, *Electro-Optical Imaging System Performance*, Norwood, MA: Artech House, 1993, p. 184.

[7] Driscoll, W., and W. Vaughn, *Handbook of Optics*, New York City, NY: McGraw-Hill, 1978, pp. 1–6.

[8] Wolfe, W., and G. Zissis, *The Infrared Handbook, Third Printing*, Environmental Research Institute of Michigan, 1987, pp. 2–77.

[9] Acetta, J., and D. Schumaker, *The Infrared and Electro-Optical Handbook*, *Volume 4*, Bellingham, WA: ERIM and SPIE, p. 31, 1993.

[10] Waldman, G., and J. Wootton, *Electro-Optical Imaging System Performance*, Norwood, MA: Artech House, p. 70, 1993.

[11] *Electro-Optical Handbook*, Lancaster, PA: Burle Industries, p. 65, 1989.

[12] *TV Performance Modeling*, Ft. Monmouth, NJ: Army Communication Command, 1991.

[13] Holst, G., *Electro-Optical Imaging System Performance*, Winter Park, FL: JCD Publishing, p. 337, 1995.

[14] Ellis, K., "Reflectance Phenomenology and Modeling Tutorial," Environmental Research Institute of Michigan, Ann Arbor, MI, 2011, World Wide Web at www.erim.com.

[15] Johnson, J., "Analysis of Image Forming Systems," *Proceedings of the Image Intensifier Symposium*, Warfare Electrical Engineering Department, U.S. Army Research and Development Laboratory, Ft. Belvior, VA, 1958, pp. 249–273.

[16] Holst, G., *Electro-Optical Imaging System Performance*, Winter Park, FL: JCD Publishing, p. 421, 1995.

[17] Rotman, R., and M. Kowalcyzyk, "Clutter Analysis for Modeling and Improving Human and Automatic Target Acquisition," *Proceedings of SPIE*, *Infrared Technology, XIX*, San Diego, CA, 1993.

[18] Schmieder, D., and M. Weathersby, "Detection Performance in Clutter with Variable Resolution," *IEEE Transactions and Electronic Systems*, Vol. 19, No. 4, 1983.

[19] Morey, B., K. Ellis, D. Perry, and K. Gleichman, *Infrared Signature Simulation of Military Targets*, Infrared Information Analysis Center, Environmental Research Institute of Michigan, ERIM, Ann Arbor, MI, 1994.

[20] Hodgkin, V. A., B. Kowalewski, S. Moyer, B. Teaney, N. Devitt, et al., "Comparison of MWIR and LWIR Signatures of Conventional and Human Targets and Rural and Urban Environments," *Military Sensing Symposium on Passive Sensors*, Orlando, FL, 2006.

[21] Widger, Jr., W. K., and M. P. Woodall, "Integration of the Plank Blackbody Radiation Function," *Bulletin of the American Meteorological Society*, Vol. 57, No. 10, 1976, pp. 1217-1219.

CHAPTER 6

Atmospherics

It is a common experience, especially for those living near a large city, to notice that one's view of the city skyline can vary dramatically from day to day depending on the weather (rain, sleet, fog, or pollution). The same factors that affect your ability to view distant objects also affect electro-optical and infrared sensors. The factors that affect the viewing of distant objects are the result of two primary phenomena: absorption and scattering. Absorption occurs when certain types of particles interact with the radiation from the object being viewed. Some key absorbers include water vapor, carbon dioxide, ozone, nitrous oxide, and carbon monoxide. Scattering is the redirection of the object's radiation by particles in the air. Examples of these particles are water droplets, snowflakes, smoke, airborne dust, pollution, and other man-made and natural aerosols.

Both absorption and scattering are strong functions of wavelength. For sensors that operate in the infrared, both absorption and scattering are important. For sensors operating in the visible waveband, scattering is the dominant phenomenon. Besides being a function of wavelength, many other variables have an influence on atmospheric transmission. These variables include temperature, altitude of sensor and target, humidity, barometric pressure, and turbulence. A pictorial representation of the atmosphere in the sensor system is shown in Figure 6.1.

With the exception of atmospheric (sometimes called "path") radiance, these atmospheric factors can reduce the amount of object radiation that is received by the sensor and produce a blurring of the image. This is especially apparent in long atmospheric paths. The extent of the impact depends on radiation wavelength. A graph of a typical atmospheric transmission is shown in Figure 6.2, depicting "windows" through the atmosphere in several spectral bands. This is one of the main reasons why the 3- to 5-μm and 8- to 12-μm spectral bands are commonly used for imaging in military applications. The transmission shown is a typical transmission, where the actual transmission changes with climate, temperature, humidity, and the like.

While the detailed theory of analyzing atmospheric transmission is quite complex, this chapter presents the basic theory along with methods for calculating transmission and its overall effect on sensor performance.

6.1 Atmospheric Components and Structure

The atmosphere is a gaseous envelope that surrounds Earth's surface. It extends for several hundred kilometers above Earth and eventually transitions into the rarefied medium of the solar system. The gas that constitutes the atmosphere, air, is made up of elements (of which there are 112) and compounds. The exact composition varies with geographic location and altitude. This is because both the atmospheric pressure and temperature change with the vertical structure of the atmosphere.

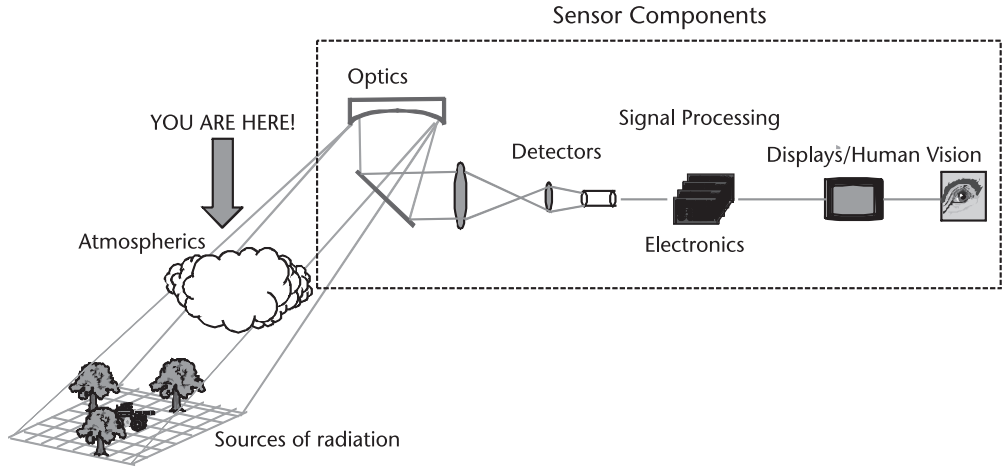


Figure 6.1 Atmospherics is the second component in the system.

On a typical day, the morning air temperature starts out cool, then gradually warms until it reaches a peak in the afternoon and finally cools down in the evening. Everyone knows that the weather (e.g., the atmospheric condition) is very unpredictable. There are warm fronts, cold fronts, isolated storms, temperature variations, and many other weather phenomena. Two separate geographic locations within the same general area can have significantly different atmospheric conditions (temperature, pressure, and composition) at the same time. This difference in atmospheric makeup affects the atmospheric transmission differently in each region. Despite these uncertainties, certain assumptions can be made allowing a reasonable prediction of the transmission of radiation through the atmosphere.

The atmosphere consists of elements and compounds. The most common elements in the atmosphere are nitrogen and oxygen, which constitute over 98% percent of the atmosphere by volume. Water vapor, carbon dioxide, nitrous oxide, carbon monoxide, and ozone are the major radiation absorbers. Water vapor has

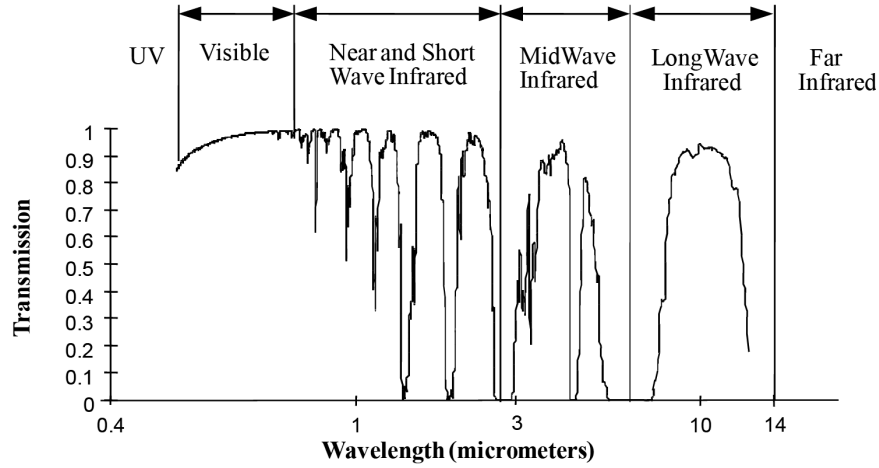


Figure 6.2 Typical atmospheric transmission for a 1-km path length.

a significant effect on absorption of infrared radiation. It is also the most variable because of evaporation of water from bodies of water, damp grounds, and condensation into clouds or dew. Carbon dioxide does not vary as dramatically as water vapor but does tend to be concentrated around large cities and heavy vegetation areas. Carbon dioxide is a strong absorber in the midwave infrared (MWIR), but the midwave band is superior for sensor applications in high humidity compared to the longwave infrared (LWIR) band.

Earth's atmosphere is generally divided into four layers, based primarily on temperature variations within each layer. These are the troposphere, stratosphere, mesosphere, and thermosphere. The temperature in the troposphere decreases with altitude at a rate of approximately 6.5 K/km. Also, between 80% and 85% of the Earth's atmospheric mass resides in the troposphere. The next atmospheric division above the troposphere is the stratosphere, where the temperature is at first a constant, then increases until approximately 50 km. The temperature increases in altitude for both the stratosphere and thermosphere. In the mesosphere, the temperature behaves in a manner similar to that of the stratosphere by decreasing in altitude to about 85 km and then strongly rising in the thermosphere. A graph of the temperature versus altitude function is shown in Figure 6.3 [1].

In the same manner as temperature, both atmospheric pressure and density vary with altitude. A tabular representation of the relationships among altitude, pressure, temperature, and density based on the U.S. Standard Atmosphere (1976) is shown in Table 6.1. These three variables—atmospheric temperature, pressure, and density—must all be taken into account when calculating atmospheric transmission. The number and behavior of the variables involved in the accurate modeling of the atmospheric effects on the flux leaving the objects of a scene can be extremely complex. However, reasonable estimates can be determined with upper and lower error bounds within assumed conditions as discussed in the following sections.

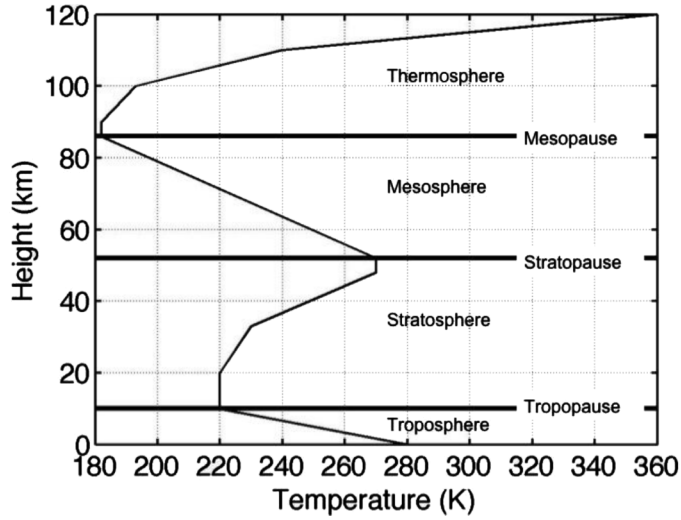


Figure 6.3 Temperature as a function of altitude.

Table 6.1 Parameters as a Function of Height

Height (km)	*Pressure (hPa)*	*Temperature (K)*	*Density (g/m^{-3})*
0.00	1.013E + 3	288	1.225E + 3
2.00	7.950E + 2	275	1.007E + 3
4.00	6.166E + 2	262	8.194E + 2
6.00	4.722E + 2	249	6.601E02
8.00	3.565E + 2	236	5.258E + 2
10.00	2.650 + 2	223	4.135E + 2
12.00	1.940E + 2	217	3.119E + 2
14.00	1.417E + 2	217	2.279E + 2
16.00	1.035 + 2	217	1.665E + 2
18.00	7.565E + 1	217	1.217E + 2
20.00	5.529E + 2	217	8.891E + 1

6.2 Atmospheric Transmission

In calculating the optical transmission from an object to a sensor, three primary processes affect the radiation: absorption, scattering, and refractive index fluctuations (turbulence). The atmospheric components (elements and compounds) discussed earlier are related to absorption and scattering, whereas the atmosphere's structural variations of temperature, pressure, and density contribute to turbulence. The effect of these three factors is both a reduction in the signal amplitude that reaches the sensor from the target and an atmospheric blurring of the image.

The first two processes, absorption and scattering, are usually grouped together under the topic of extinction. Extinction is defined as the reduction or attenuation in the amount of radiation passing through the atmosphere. Absorption is when a photon of radiation is absorbed by a gaseous molecule of the atmosphere that converts the photon into kinetic energy, thereby changing the energy level of the molecule or its component atoms and electrons. This energy change of the gaseous molecule manifests itself as a temperature change. When radiation is scattered, the direction of the incident radiation is changed by collision (and reflection) of a photon with an atmospheric molecule or particle. This process can be thought of as the atmospheric molecule momentarily capturing the incident radiation and then sending it out unchanged, but in all directions. The amount of radiation dispersed in a particular direction depends on the properties of the air particle interacting with the incident radiation.

To understand how the extinction affects the transmission of radiation through the atmosphere, visualize a single wavelength of incident radiation passing through an infinitesimal distance dx. The change in flux is

$$d\Phi = -\sigma\Phi\, dx \quad [\mathrm{W}] \tag{6.1}$$

where the incident radiation is represented by Φ and σ is the extinction coefficient as shown in Figure 6.4. The negative sign indicates the reduction of power, or ra-

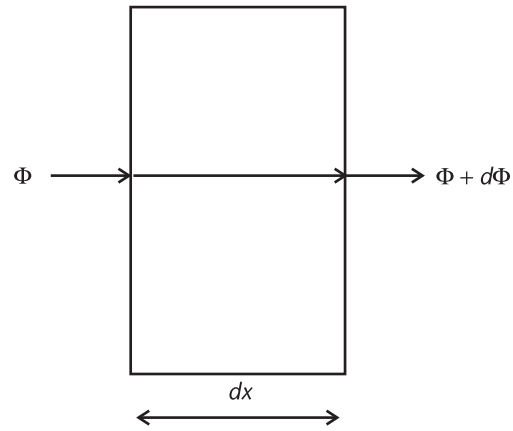


Figure 6.4 Transmission through an infinitesimal distance of atmosphere.

diation, from the absorption and scattering processes. In general, the extinction coefficient is made up of two components:

$$\sigma = a + \gamma \quad [\text{km}^{-1}] \tag{6.2}$$

where a is the absorption coefficient and γ is the scattering coefficient. Both the absorption and scattering coefficients depend on the incident wavelength. Because of the interaction of the incident radiation with molecules comprising the medium, the radiation at the output is changed to $\Phi + d\Phi$. Equation (6.1) can be integrated to give the attenuation for a finite distance through a *homogeneous medium*:

$$\Phi = \Phi_o e^{-\sigma x} = \Phi_o \tau \quad [\text{W}] \tag{6.3}$$

where $\tau = e^{-\sigma x}$ is the transmittance of the atmosphere over a distance x. This principle is known as the Beer-Lambert law (sometimes shortened to Beer's law). As mentioned above, the extinction coefficient is highly dependent on wavelength, so Beer's law is usually written as

$$\tau(\lambda) = e^{-\sigma(\lambda)x} \quad [\text{unitless}] \tag{6.4}$$

Example 6.1
The transmittance of air $\tau_{atm}(\lambda)$ is the product of the molecular extinction transmittance and the aerosol extinction transmittance:

$$\tau_{atm}(\lambda) = \tau_m(\lambda)\tau_a(\lambda) \tag{6.5}$$

A CO_2 laser with a 10.6-μm wavelength is propagated through a 2-km horizontal path length. Determine the molecular transmittance given that the molecular extinction is 0.385 per kilometer (defined for the given wavelength and relative humidity). The solution gives the transmittance at

$$\tau_m(\lambda) = e^{-(0.383)(2\,\text{km})} = 0.465 \quad \text{[unitless]} \tag{6.6}$$

If the aerosol extinction is negligible (i.e., its transmittance is unity), a 10W laser provides 4.65W at a distance of 2 km.

6.3 Absorption

Absorption is fundamentally a quantum process in which an atmospheric molecule absorbs the energy from some incident photon. Absorption by an atmospheric molecule changes its internal state, increasing its energy, and resulting in a temperature change. When discussing absorption, it is best to consider the quantum nature of the atmospheric molecule. From introductory physics, the energy structure of a molecule is illustrated in Figure 6.5. The energy gap E_g required for absorption is given in Einstein's equation, $E = h\nu$. where E is the energy of the photon, h is Planck's constant (6.626×10^{-14} J-s), and ν is the frequency of the incident light.

The energy gap E_g must be exceeded by a photon for an electron to change energy states within a molecule. There are a large number of "allowed" energy state changes within a molecular energy structure, so its absorption properties contain a spectral dependence. For a photon to be absorbed by a molecule, it must be at a sufficient wavelength such that $E_{photon} \geq E_g$ holds true. This energy difference is related to the different rotational and vibrational states of the particular atmospheric molecules. If the appropriate molecular energy band does not exist, the corresponding photons at that particular wavelength will not interact (to the first order) with that particular atmospheric molecule. Because of the quantum mechanical makeup of atmospheric gases, there is no significant absorption or radiation in the visible wavelengths (except for H_2O absorption between 0.65 and 0.85 μm). While in the infrared region of the spectrum, a number of the molecules present are highly absorptive. Specific atmospheric components that are absorptive in the infrared include the diatomic molecules nitrogen and oxygen; water; and carbon dioxide. Plots of the absorption characteristics for the primary components versus wavelength are shown in Figure 6.6.

Example 6.2
A molecule is sensitive [2] to photons with energies equal to, or slightly greater than, its molecular bandgaps. Lower energy photons pass and are not absorbed. Determine the

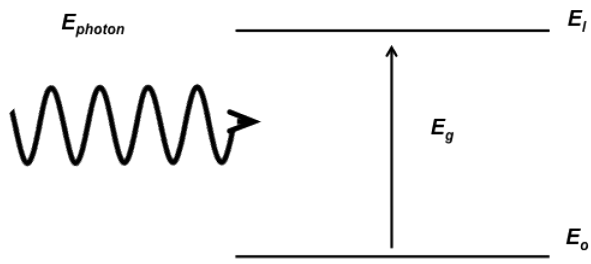


Figure 6.5 Absorption of a photon.

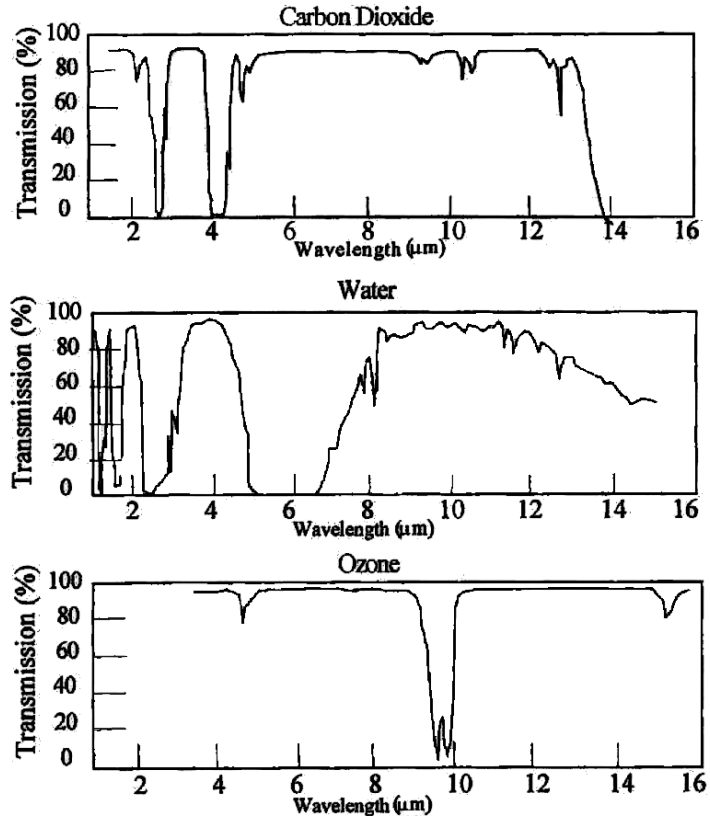


Figure 6.6 Transmission of atmospheric gases for a 1-km path length.

wavelength with the most sensitivity (absorption) for a molecule with a gap at 0.18 eV. Planck's constant comes in two forms: 6.63×10^{-34} J-s or 4.14×10^{-15} eV. The speed of light is 3.00×10^{8} m/s. The solution is determined by Einstein's relationship

$$E_g = h\nu \quad [\text{eV}] \tag{6.7}$$

With E at 0.18 eV, the frequency is 4.35×10^{-13}/s, or Hertz. The wavelength can then be determined from

$$\lambda = \frac{c}{\nu} \quad [\text{m}] \tag{6.8}$$

yielding a wavelength of 6.9 μm.

The spectral graphs shown in Figure 6.6 are low-spectral-resolution graphs, where individual absorption lines cannot be seen (i.e., transmission has been averaged over some small wavebands). High-resolution-transmission graphs (1/16 wavenumber) are shown in Figure 6.7 for a 1-km horizontal path at sea level. Frequently, transmission is plotted as a function of wavenumber, where the wavenumber is the inverse wavelength $1/\lambda$ and wavelength is given in centimeters. The graphs are produced by FASCODE, a high-resolution atmospheric spectral analysis program by Ontar.

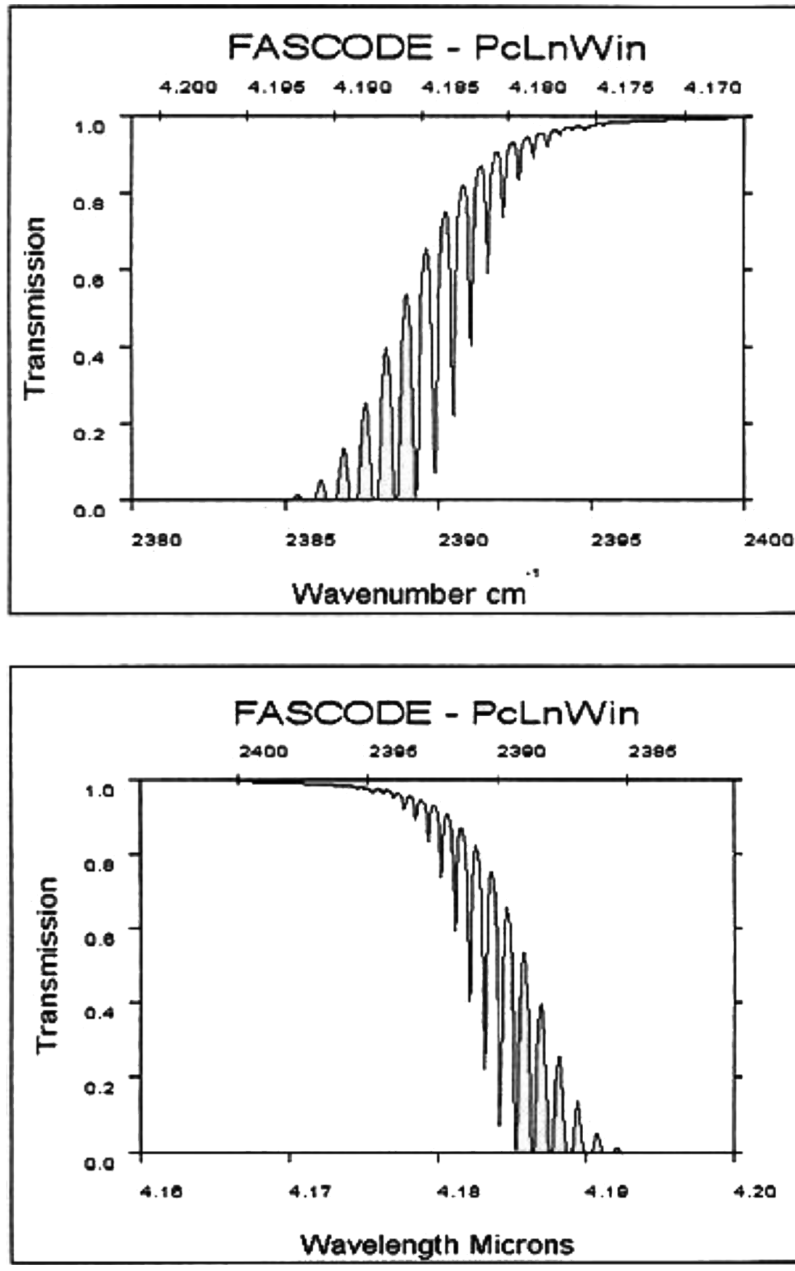


Figure 6.7 High-resolution atmospheric transmission. (Courtesy of J. Schroeder, Ontar.)

6.4 Scattering

The process of scattering is the result of photons colliding with atmospheric particles whereby the photon energy is reradiated in all directions. The fraction of the photon's energy extracted and the angular pattern that it is reradiated are a function of the relative atmospheric particle size to that of the incident photon wavelength. Scattering differs from absorption because the radiant energy scattered remains in the form of radiation, but is redistributed from the change in the direction of propa-

gation of the incident radiation. Typical angular scattering patterns are shown in Figure 6.8.

In general, the scattering models are divided into three categories. These categories are based on the relationship between the particle size and the wavelength of the incident photon. These models are the Rayleigh scattering, Mie scattering, and geometric optics models, which are discussed below. Typical examples of some atmospheric particles are shown in Table 6.2.

The first type of scattering is Raleigh scattering. It is used when the particle radius is approximately less than one-tenth of a wavelength. The scattering coefficient is proportional to λ^{-4}. As shown in Figure 6.8, the angular distribution is symmetrical. It is Rayleigh scattering that leads to the blue color of the sky. The blue wavelengths have a strong scattering interaction with the atmosphere. The blue is scattered in all directions so that the sky appears blue, whereas the red end of the visible spectrum provides little scattering. The attenuation of light for Rayleigh scattering can be represented by the first-order approximation, given by simplification of (6.3):

$$\Phi(x) = \Phi_o e^{-\gamma x} \quad [\mathrm{W}] \tag{6.9}$$

where only the scattering coefficient γ is considered. The Rayleigh scattering coefficient is given by

$$\gamma = \gamma_r N \quad [\mathrm{km}^{-1}] \tag{6.10}$$

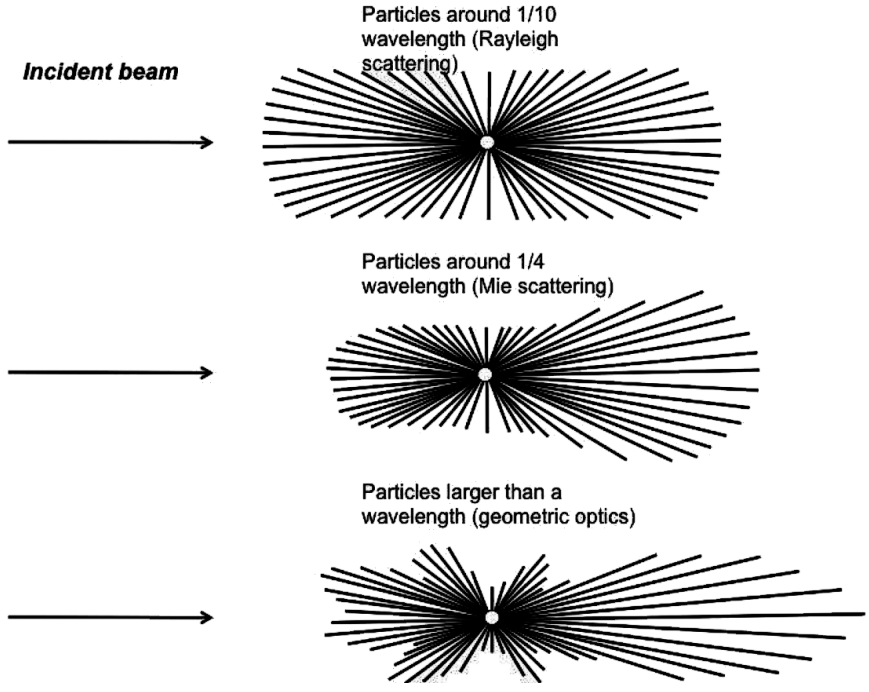


Figure 6.8 Scattering by particles. (Adapted from Waldman and Wooton [3].)

Table 6.2 Atmospheric Particles

Particle Type	*Radius (μm)*	*Density (per cm³)*
Air molecules (Rayleigh)	10^{-4}	10^{19}
Haze particles (Rayleigh)	10^{-2}–1	10–10^{3}
Fog droplet (Mie)	1–10	10–100
Raindrops (Geometric)	10^{2}–10^{4}	10^{-5}–10^{-2}

where γ_r is the Rayleigh scattering cross section and N is the molecular number density. Examples of the molecular number density are shown in Table 6.2. The Raleigh scattering cross section is given by

$$\gamma_r = \frac{8\pi^3(n^2-1)^2}{3N^2\lambda^4} \quad [\text{km}^2] \tag{6.11}$$

where n is the index of refraction. Rayleigh scattering dominates more in the UV and EO wavelengths since the Rayleigh cross section is proportional to λ^{-4}. For a 0.5-μm wavelength and good weather, the Rayleigh scattering coefficient is about 0.0172 km^{-1} [4].

The second type of scattering to be discussed is Mie scattering or aerosol scattering. This occurs when the wavelength is approximately the same size as the particle. Detailed Mie theory is based on classical electromagnetic equations with continuity conditions of the boundary between the particles and its surroundings. Most atmospheric models make certain assumptions to simplify the theory. One such assumption is that the atmospheric particles can be represented by simple shapes such as cylinders, ellipsoids, and spheres. It is beyond the scope of this book to derive the details of Mie theory. Instead, a first-order approximation is provided. This type of scattering is concentrated in the forward direction as shown in Figure 6.8. The attenuation given by Mie scattering is of the same form as Raleigh scattering and is expressed by

$$\Phi(x) = \Phi_o e^{-\gamma_m x} \quad [\text{W}] \tag{6.12}$$

where γ_m is the aerosol attenuation coefficient. The aerosol attenuation coefficient is a function of the aerosol density and is related by

$$\gamma_m(h) = \frac{M(h)}{M(0)}\gamma_m(0) \quad [\text{km}^{-1}] \tag{6.13}$$

where $M(h)$ denotes the aerosol density at a height h, and $\gamma_m(0)$ is the aerosol coefficient at sea level (density at sea level ~200 cm^{-3}). The aerosol coefficient also varies with wavelength. Several examples of the aerosol attenuation coefficient are given in Table 6.3 [5].

The third type of scattering occurs when the particles are much larger than the wavelength. An example of this scattering type is that of raindrops. For this case, the light is primarily scattered in the forward direction. This case can be approximated with geometrical optics.

Table 6.3 Aerosol Attenuation Coefficients

Wavelength (μm)	*Aerosol Coefficient at Sea Level*
0.30	0.26
0.40	0.20
0.50	0.167
0.60	0.150
0.70	0.135
0.80	0.127
0.90	0.120

6.5 Path Radiance

Path radiance is the natural emission of radiation by atmospheric particles within a sensor FOV. Given a path length between a source and a sensor, the atmospheric particles in this path contribute radiation to the total flux on the detector. To fully describe the flux on the detector, the path radiance must be taken into consideration.

Path radiance can be described in a number of different ways, but the simplest method is to model the particles as blackbodies. If the atmospheric particles are considered a blackbody at a particular temperature $L_{BB}(T_a)$, then the amount of path radiance seen by the sensor is

$$L_{path}(\lambda) = \left[1 - t_{path}(\lambda)\right] L_{BB}(t_a) \quad [\text{W/cm}^2\text{-sr-mm}] \tag{6.14}$$

where $t_{path}(\gamma)$ is the atmospheric transmission of the path. Given any sensor viewing geometry where the source fills the detector (i.e., the image of the object is larger than the detector), the total flux on the detector can be calculated using an effective total radiance:

$$L_{total}(\lambda) = L_{source}(\lambda) t_{path}(\lambda) + L_{path}(\lambda) \quad [\text{W/cm}^2\text{-sr-mm}] \tag{6.15}$$

If the source does not fill the detector and the background contributes flux on the detector, then background radiance must also be added to the total radiance.

There are many arguments on when to include path radiance and when to ignore path radiance. From (6.14), the path radiance becomes small when the path transmission is high. With a large transmission and cold-air particles (compared with the source), the path radiance may be negligible. However, a small path transmission, high air temperature (with respect to source), or both can provide significant path radiance. An additional note is that large path transmission corresponds to small paths, so path radiance may be negligible for short ranges.

6.6 Turbulence

Turbulence is a term used to describe time-varying temperature inhomogeneities of the atmosphere. The temperature (and pressure) variations result in index of

refraction inhomogeneities. The index of refraction variations cause the direction of light propagation to "bend" in various directions. These directional changes vary with time. Turbulence is responsible for a variety of effects, such as temporal intensity fluctuations (i.e., scintillations). An example of scintillations is the twinkling of the stars (a small source through a large atmospheric path length). Turbulence affects imaging not only by visible sensors but also by infrared sensors. An example of infrared turbulence can be considered when viewing a distant object across a parking lot in the middle of the summer. The sun heats the pavement, causing the air above it to become heated. This heated air is unstable and radiation passing through it deviates from its original path. This phenomenon causes image motion and blurs the image. It is especially noticeable when viewing small, distant objects. It differs from extinction since motion or blurring is not uniform across the image. Overall, extinction affects the entire image in a uniform manner.

As with other sections of this chapter, a complete development of turbulence is out of scope, hence the focus here is on several key parameters. These parameters are defined along the optical path for small-scale atmosphere turbulence. Large-scale atmosphere effects are not generally considered in atmospheric modeling for imaging sensors.

Turbulence consists of time-varying fluctuations that change the localized index of refraction. The fluctuations of the refractive index in the atmosphere can be described by the index structure function:

$$Dn(r) = < (n_1 - n_2)^2 > \tag{6.16}$$

where n_1 and n_2 are the indices of refraction at two points separated by a distance r. The notation < > describes the ensemble average. In a similar manner, a temperature structure function is defined by

$$D_T(r) = < (T_1 - T_2)^2 > \tag{6.17}$$

where T_1 and T_2 are the temperatures at two points separated by distance r. The Kolmogorov-Obukhov law states that differences in indices and temperatures are proportional to the distance to the two-thirds power:

$$D_n(r) = C_n^2 r^{2/3} \tag{6.18}$$

and

$$D_T(r) = C_T^2 r^{2/3} \tag{6.19}$$

where C_n^2 and C_T^2 are constants called the *index structure parameter* and the *temperature structure parameter*, respectively.

Solving for the temperature structure parameter using (6.17) and (6.19) shows that the temperature structure parameter can be measured directly:

$$C_T^2(r) = \frac{< (T_1 - T_2)^2 >}{r^{2/3}} \tag{6.20}$$

The two-thirds power rule holds for many cases. It describes the heat transfer between the surface and the atmosphere. The two-thirds power functions are valid for stable conditions (night), neutral conditions (sunrise and sunset), and for some daytime conditions where there is warm air over cold ground (e.g., snow or ice on the ground). For unstable conditions (free convection over hot surfaces during the day), a four-thirds power is more appropriate.

The index structure parameter is related to the temperature structure parameter by

$$C_n^2 = \left|\frac{\partial n(\lambda)}{\partial T}\right|^2 C_T^2 \quad [\mathrm{m}^{-2/3}] \tag{6.21}$$

For dry air and optical wavelengths, (6.21) can be approximated by [6]

$$C_n^2 \approx \left[79\times10^{-6}\frac{P}{T}\right]^2 C_T^2 \quad [\mathrm{m}^{-2/3}] \tag{6.22}$$

where P is the atmospheric pressure in millibars and T is the temperature in Kelvin. The index structure parameter ranges from 1×10^{-15} $\mathrm{m}^{-2/3}$ for weak turbulence to 5×10^{-13} $\mathrm{m}^{-2/3}$ for strong turbulence. Factors that increase the index structure parameter are strong solar heating, very dry grounds, clear nights with little wind, low altitudes, and surface roughness. Factors that provide reductions in the index structure parameter are heavy overcast, wet surfaces, high winds, and high altitudes.

It is useful to estimate C_n^2 as a function of weather conditions. Computer programs such as IMTURB and PROTURB have been developed by the U.S. Army Atmospheric Sciences Lab for these types of estimations. Also, a simple model [7, 8] has been developed using data taken by NVESD. It is based on meteorological parameters and the Talmudic concept of temporal hours.

Talmudic hours are not 60 min, but are 1/12th of the time between sunrise and sunset. In the winter, temporal hours are shorter than 60 min and in the summer they are longer than 60 min. An example is that 3 temporal hours corresponds to one-quarter of the daylight time.

A weighting factor is used in the estimate of C_n^2 and is related to the temporal hour. The weighting factor is given in Table 6.4. A regression model was performed on C_n^2 as a function of this weighting factor, the temperature, relative humidity, and wind speed. Based on the temporal hour and this meteorological data, the index structure parameter in $\mathrm{m}^{-2/3}$ can be estimated by

$$\begin{aligned} C_n^2 = {} & 3.8\times10^{-14}W + 2.0\times10^{-15}T - 2.8\times10^{-15}RH + 2.9\times10^{-17}RH^2 \\ & - 1.1\times10^{-19}RH^3 - 2.5\times10^{-15}WS + 1.2\times10^{-15}WS^2 \\ & - 8.5\times10^{-17}WS^3 - 5.3\times10^{-13} \end{aligned} \tag{6.23}$$

where W is the weight given for the temporal hour in Table 6.4, T is the air temperature in Kelvin, RH is the relative humidity in percent, and WS is the wind speed in meters per second.

Table 6.4 Temporal Hour and Index Structure parameter weight

Event	*Temporal Hour Interval*	*Relative Weight (W)*
Night	... until –4	0.11
Night	–4 to –3	0.11
Night	–3 to –2	0.07
Night	–2 to –1	0.08
Night	–1 to 0	0.06
Sunrise	0 to 1	0.05
Day	1 to 2	0.1
Day	2 to 3	0.51
Day	3 to 4	0.75
Day	4 to 5	0.95
Day	5 to 6	1.00
Day	6 to 7	0.9
Day	7 to 8	0.8
Day	8 to 9	0.59
Day	9 to 10	0.32
Day	10 to 11	0.22
Sunset	11 to 12	0.1
Night	12 to 13	0.08
Night	over 13	0.13

This model has been validated over desert sand as well as vegetation surfaces. It is valid for temperatures of 9° to 23°C, relative humidities of 14% to 92%, and wind speeds of up to 10 m/s. The model applies to index structure parameters at heights of up to 15m. For heights that are not near Earth's surface (beyond 15m), C_n^2 can be scaled according to the Tatarski [9] model:

$$C_n^2(h) = C_n^2(0)h^{-4/3} \tag{6.24}$$

where h is the height in meters and $C_n^2(0)$ is the ground-based index structure parameter. A modification to this model is provided by Sadot et al. [8] for cases that include significant aerosol effects.

The effect of turbulence on EO systems is most pronounced for coherent systems (lasers), where the propagation of the optical wavefront is most critical. Strong turbulence also affects incoherent imaging systems by image blurring and smearing with the loss of high-frequency information. The effects of turbulence on imaging through the atmosphere can be characterized by atmospheric MTFs. While good progress has been seen in the modeling of atmospheric turbulence, there are no accepted standard turbulence models.

Example 6.3

Determine the index structure parameter for midday at a temperature of 27°C, relative humidity of 40%, and a wind speed of 2 m/s. The solution is determined by using the following parameters in (6.23):

$$W = 1.0$$
$$T = 300\text{K } (27°\text{C})$$
$$RH = 40\%$$
$$WS = 2 \text{ m/s}$$

These parameters give an index structure parameter of 3.5×10^{-14} $m^{-2/3}$.

Example 6.4
Determine the index structure parameter for a height of 50m given that the ground-based index structure parameter is 2×10^{-14} $m^{-2/3}$. The solution is determined using the Tatarski model of (6.24):

$$C_n^2(50) = 2.0 \times 10^{-14}(50)^{-\frac{4}{3}} = 1.09 \times 10^{-16}\,\text{m}^{-2/3}$$

Note that the index structure parameter decreases significantly for a height of only 50m.

6.7 Atmospheric MTF

In the previous sections, various atmospheric phenomena that affect optical radiation were reviewed. Absorption and scattering can be viewed as a reduction in the amount of radiation that reaches a sensor, while scattering and turbulence result in image blurring and loss of detail. The blurring is quantified to describe its overall degradation effect on sensor performance. This degradation is characterized in terms of an atmospheric MTF. Atmospheric MTF can roughly be described as a reduction in contrast as a function of spatial frequency. The atmospheric MTF can be divided into two MTFs: an aerosol MTF and a turbulence MTF. This approach is not exact for atmospheric modeling but provides a first-order approximation. A limitation to this approach exists because MTF theory is based on linear invariant processes. Turbulence is not necessarily uniform across an image but is often assumed so for modeling purposes.

The atmospheric turbulence MTF has two forms depending on the exposure time. For the long-exposure-time case, the MTF is given by

$$MTF_{le}(\xi) = \exp\left[-57.53\xi^{\frac{5}{3}} C_n^2 \lambda^{-\frac{1}{3}} R\right] \tag{6.25}$$

where ξ is the spatial frequency in cycles per radian, R is the path length in meters, and λ is the wavelength in meters. If R is given in kilometers, the ξ can be plotted in milliradians. The term *short exposure time* is defined by Goodman [10] as the "time required to freeze the atmospheric degradation and thus eliminate the time-averaging effects." This depends on wind velocity but is generally less than 0.01 to 0.001 sec. The short-exposure-time turbulence MTF is given by

$$MTF_{se}(\xi) = \exp\left\{-57.53\xi^{\frac{5}{3}}C_n^2\lambda^{-\frac{1}{3}}R\left[1-\mu\left(\frac{\lambda\xi}{D}\right)^{\frac{1}{3}}\right]\right\} \tag{6.26}$$

where μ is 1 for the nearfield and 0.5 for the farfield, and D is the sensor's entrance pupil diameter. The nearfield is defined when $D >> (\lambda R)^{1/2}$ and the farfield when $<< (\lambda R)^{1/2}$.

Example 6.5

Determine the MTF for a long-exposure-time sensor that operates in the longwave over a horizontal ground path length of 8 km. The atmospheric conditions are such that the index structure parameter is 1×10^{-14} $m^{-2/3}$. The solution is given by the long exposure MTF of (6.25). Because the sensor is a longwave sensor, a wavelength of 10 μm is assumed. The MTF is shown in Figure 6.9. Note that while (6.25) was developed for units of cycles per radian, the MTF of Figure 6.9 has been plotted against cycles per milliradian.

The aerosol MTF is a strong function of sensor FOV, sensitivity, and limiting bandwidth. The aerosol MTF comprises two components: a low-spatial-frequency function and a high-spatial-frequency function. The aerosol MTF is approximated [11] by

$$MTF_a(\xi) = \begin{cases} \exp\left[-A_aR - S_aR\left(\dfrac{\xi}{\xi_c}\right)^2\right] & \text{for} \quad \xi \le \xi_c \\ \exp\left[-(A_a + S_a)R\right] & \text{for} \quad \xi > \xi_c \end{cases} \tag{6.27}$$

where ξ_c is the cutoff spatial frequency that is approximately a/λ, a is the particulate radius, A_a is the atmospheric absorption coefficient, and S_a is the atmospheric scattering coefficient. These quantities are usually measured, but Sadot and Kopeika [11] have outlined a detailed method for calculating these parameters. Holst [12] states

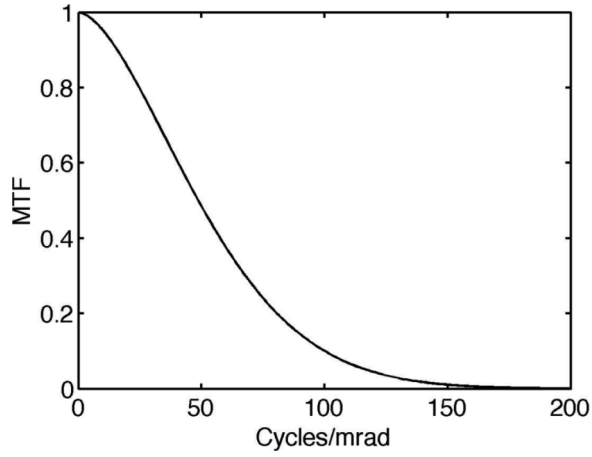


Figure 6.9 Turbulence MTF.

that aerosol contributions to atmospheric MTF are significant when the target radiance scattering approaches the radiance scattering attributed to the background. For small signal targets such as tanks in tactical background scenarios in the infrared, the target scattered signal is small compared with the background scattered signal, and aerosol MTF can be neglected. For significant target scattering such as rocket plume signals against sky backgrounds in EO wavelengths, aerosols must be considered.

6.8 Models

An accurate atmospheric transmission calculation requires that all molecular, aerosol, and precipitation effects be taken into account. When exact results are required, such as the design and manufacturing of a sensor system, a detailed model must be used. Three such models are LOWTRAN, FASCODE, and MODTRAN. These models are widely used in both government and private industry. The intent here is to make the reader aware of these models and their basic functionality.

Until recently, LOWTRAN was the most widely used and accepted model for atmospheric transmittance. The original computer code was developed at the Air Force Geophysics Laboratory in Cambridge, Massachusetts. The commercial software was developed by Ontar Corporation. This model was broken down into two components: absorption and scattering. The resolution of the LOWTRAN model in wave numbers ($1/\lambda$) is 20 cm^{-1}. The code calculates transmittance from 0.25 to 28.5 μm. Four example curves generated with LOWTRAN are shown in Figure 6.10. The curves shown are the transmission of a 1-km horizontal atmospheric path length for the climates given. It is readily apparent that water plays a large part in the decrease in LWIR transmission. In the cooler climates, the absolute humidity is much lower, so the water absorption is less. In the warmer climates, the air can hold more gaseous water, so the transmission is less. The water does not affect the visible or midwave as much as it does the longwave. This can be verified in LOWTRAN with the water absorption values. Also, note that the three sensor bands addressed in this text (electro-optical, midwave, and longwave) have atmospheric windows where the transmission is good. Also, longer path lengths (e.g., 10 km) would produce a much more severe transmission curve.

MODTRAN is currently the most widely used and accepted model for atmospheric transmittance. It evolved from LOWTRAN and computes both transmission and radiance at 2-cm^{-1} wavenumber resolution. An example of MODTRAN transmission is shown in Figure 6.11. The transmission shown is for a 1-km horizontal path at sea level. FASCODE is a line-by-line model that computes the transmission of each absorption line using the spectroscopic parameters in the HITRAN database. MODTRAN uses a two-parameter band model, derived from HITRAN, as its database of absorption parameters. Both models have the capability to handle arbitrary geometries and view conditions, and allow the user to select from several model atmospheres, aerosol profiles, and other weather conditions.

The commercial versions of MODTRAN and FASCODE are available from the Ontar Corporation (http://www.Ontar.com). More information about the HITRAN database can be obtained from the http://www.HITRAN.com website.

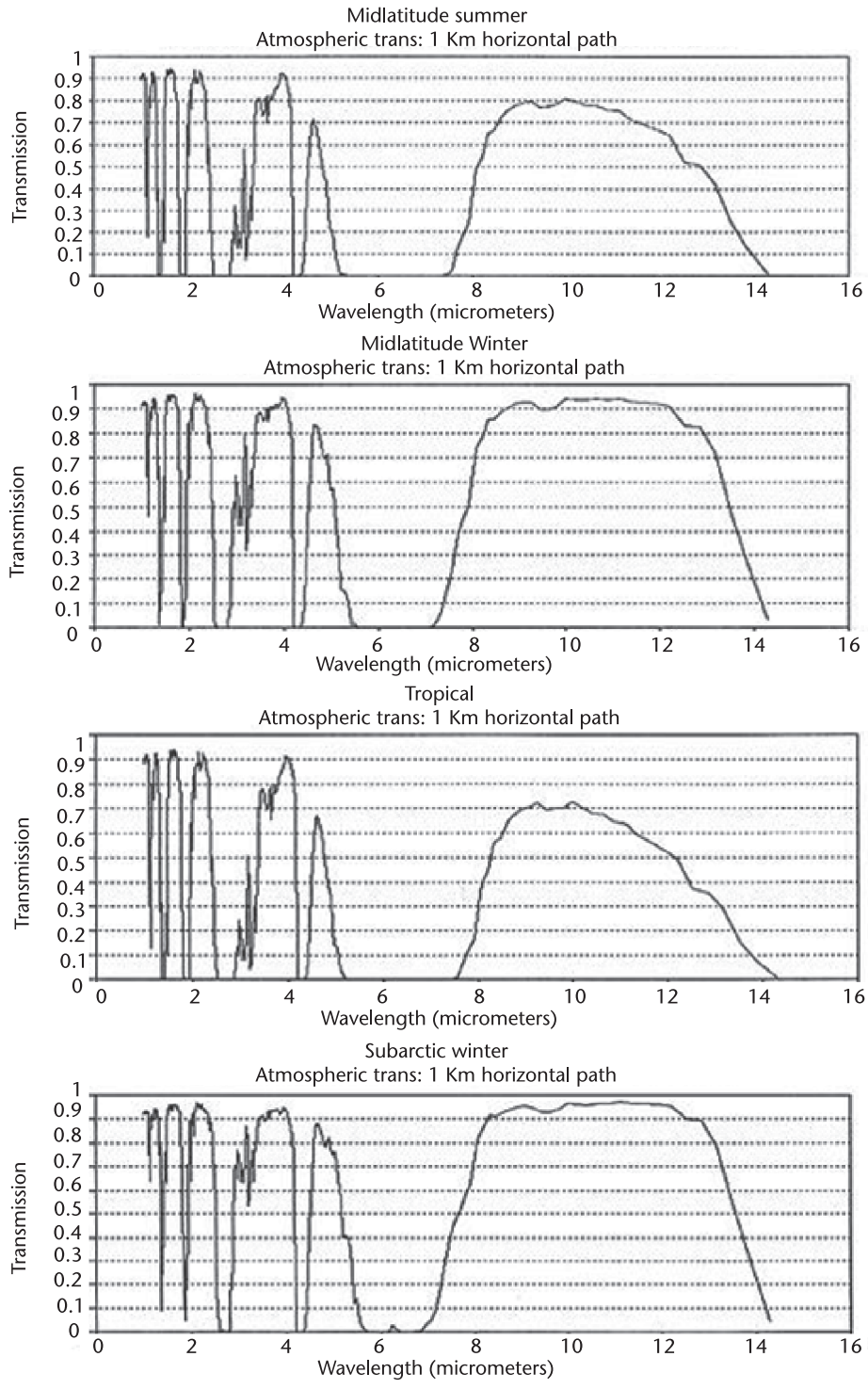


Figure 6.10 LOWTRAN transmission curves.

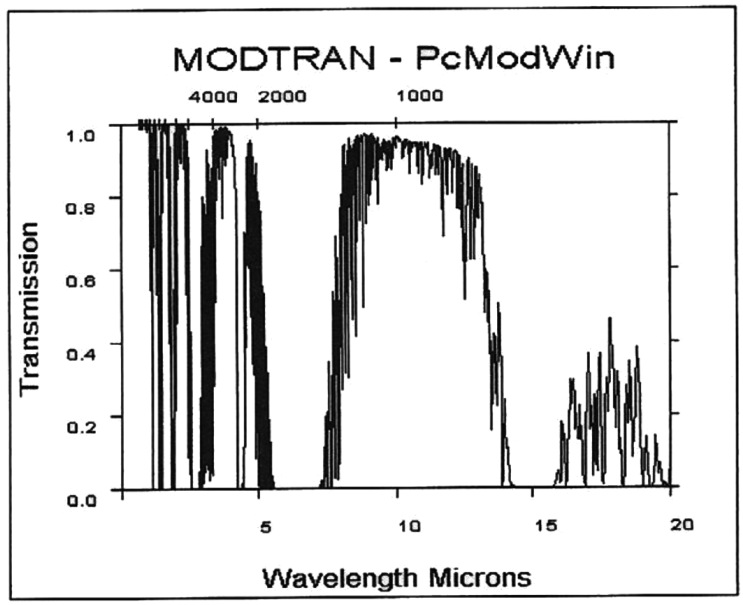


Figure 6.11 MODTRAN results. (Courtesy of J. Schroeder, Ontar Corp.)

6.9 Model Discussion

To determine an exact amount of radiation originating from an object and propagating to the sensor detector (i.e., less path radiance and background radiance), we must consider the atmospheric transmission as a function of wavelength. The radiation received by the detector within a sensor can be expressed by

$$P_{sensor} = \int_{\lambda_1}^{\lambda_2} \frac{M_{source}(\lambda)}{\pi} A_{source} \Omega_{sensor} \tau_{atm}(\lambda) \tau_{sensor}(\lambda) d\lambda \tag{6.28}$$

where

- $\frac{M_{source}(\lambda)}{\pi}$ = source (Lambertian) radiance (W/cm^2-sr-μm)
- A_{source} = source area seen within the sensor FOV (cm^2)
- Ω_{sensor} = solid angle of the sensor entrance pupil as seen by the source (sr)
- $\tau_{atm}(\lambda)$ = transmission of the atmosphere between the source and the sensor (unitless)
- $\tau_{sensor}(\lambda)$ = the transmission of the sensor optical path (unitless).

Once the wavelength integration of the object radiance, atmospheric transmission, and sensor transmission is performed, there is no method to separate source distribution and atmospheric transmission, with two exceptions: (1) The source exitance is constant within the spectral band, or (2) the atmospheric transmission is constant over the spectral band.

A common, but incorrect, practice is to perform a band transmission measurement of the atmosphere by viewing a near blackbody source and then to view a distant blackbody source with a broadband sensor that matches the band of interest. The blackbodies are both at the same temperature to provide a similar signal, and the collection geometries are accounted for. The ratio of the near power to the distant power is considered the average atmospheric band transmission. Even assuming the sensor transmission is constant and taking the constants outside the (6.28) integral, this approximation gives

$$\tau_{band} = \frac{c_1 \int_{\lambda_1}^{\lambda_2} M(\lambda)\tau_{atm}(\lambda)d\lambda}{c_2 \int_{\lambda_1}^{\lambda_2} M(\lambda)d\lambda} \tag{6.29}$$

This measurement gives a weighted transmission that is source dependent. An average transmission that is independent of source distribution cannot be taken from (6.29). However, the assumption that source distribution is relatively constant for low-temperature blackbody sources over the 8- to 12-μm spectral band has been known to give some reasonable error bounds. This is not the case for EO sources and 3- to 5-μm sensor bands. A second method for determining average band transmission has become popular with the spectral radiometer. The spectral radiometer measures the power on the detector [i.e., (6.28)] in fine spectral slices such that the ratio of the near- and far- detector responses can be normalized over wavelength to give $\tau_{atm}(\lambda)$. The average band transmission is then computed as follows:

$$\tau_{band} = \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \tau_{atm}(\lambda)d\lambda \tag{6.30}$$

Although this technique gives a non–source-weighted atmospheric transmission over a band of interest, it still does not describe the detector response through the atmosphere for a given source. Again, (6.28) describes the detector response that is a function of a spectral atmospheric transmission and source-radiance product. This response is not equal to the average source radiance and average atmospheric transmission product.

Now that average atmospheric band transmission quantities have been presented in a negative light, it is time to realize that we do not live in a perfect world. One does not always have the luxury of working with spectral quantities, and instruments for spectral measurements are not always available. It is a good idea when it is necessary to work with average values to quantify the errors associated with particular assumptions. While the approximations above may provide acceptable errors in the LWIR band, they may not in the EO and MWIR bands. Either way, quantification of error bounds is always a good sanity check.

Atmospheric transmission is the subject of many research projects and our knowledge is continually becoming more refined. MODTRAN, LOWTRAN, and other models are being updated frequently to account for more atmospheric phenomena. The data presented in this chapter provide merely a brief introduction to the issues of atmospheric transmission applied to I^2R and EO systems. While simple models are useful for back-of-the-envelope analyses, more exact system responses can be accomplished with sophisticated computer models.

6.10 Some Practical Considerations

A frequent argument in infrared imaging is whether to use MWIR or LWIR systems given a particular application. Historically, LWIR systems have been used heavily by world armies and MWIR systems have been used heavily by world navies. Today, this still holds true and the graph in Figure 6.12 from [15] provides a partial explanation. Many factors must be considered when selecting a band, including resolution and target contrast. Another factor is atmospheric transmission. In Figure 6.12, a source target contrast of 2K is assumed in both the MWIR and LWIR. This target contrast is at a range of 0 km and it is identical for the MWIR and LWIR bands (which is frequently the case for real targets). As the range between the sensor and the target increases, the apparent target contrast decreases as shown. The numbers next to the points in the graph are the range between the sensor and the target in kilometers. Three atmospheres are shown: a U.S. standard atmosphere, a desert atmosphere, and a tropical/maritime atmosphere. The desert

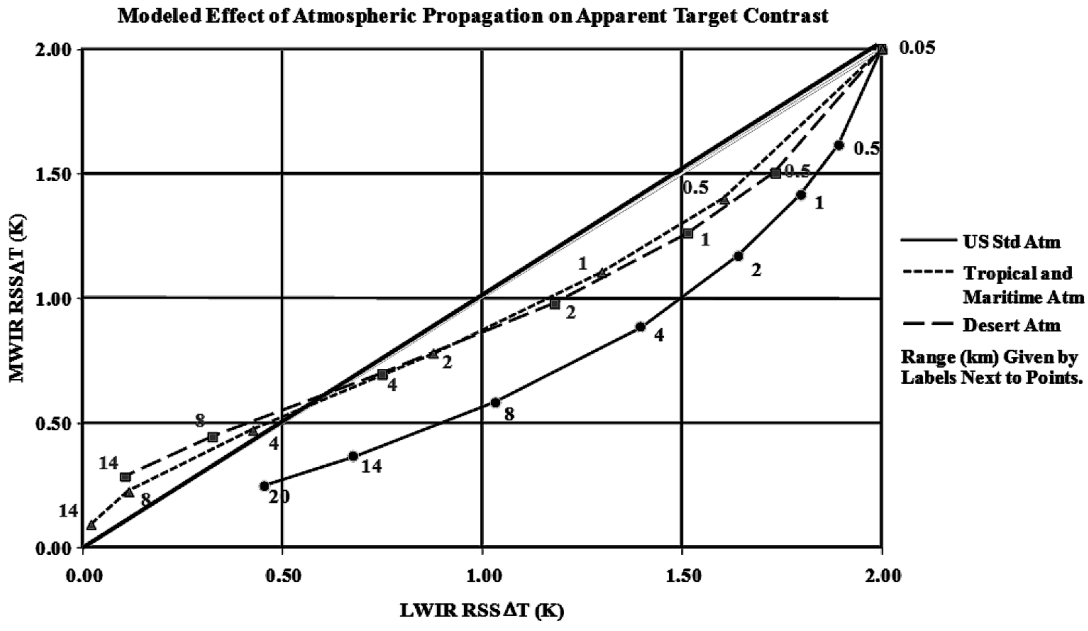


Figure 6.12 Target contrast as a function of range in the atmosphere.

is not an extremely dry desert, but one that is not too far from the Persian Gulf. If the points are above the diagonal line, the MWIR target contrast is higher; if the points are below the diagonal line, the LWIR target contrast is higher. It is readily apparent that the LWIR contrast is always higher in a land-based, U.S. standard–like atmosphere. This partially explains why armies choose LWIR sensors for military applications. For the atmospheres with more water in the air, such as tropical/maritime and desert, there is a crossover from LWIR to MWIR dominant contrast somewhere around 4 km. As the range increases and the apparent target contrast moves below 0.5K, MWIR sensing becomes more important. This contrast is one of the reasons that world navies use MWIR sensors more than LWIR sensors on military platforms.

Another practical consideration relates to when a sensor is "turbulence limited"; that is, when the atmospheric blur is larger than the sensor blur. Figure 6.13 shows bar targets imaged through a significant amount of turbulence. The set of bar targets on the right is at a closer range than the set of bar targets on the left. The blur in the image is mostly due to turbulence, and a distortion occurs as well. The blur is not shift invariant (sometimes called non-isoplanatic by atmospheric researchers) and the MTFs presented earlier are the average MTFs over space in the image. It is important to realize that the blur due to turbulence depends on range, so the MTF varies with range. The MTF becomes more band limited at longer ranges. Therefore, in any system range calculation, the MTF calculation is performed at each range (unlike the other MTFs associated with optics, detectors, electronics, and displays).

Figure 6.14 is Weiss-Wrana data from [16] that describes the C_n^2 as a function of time of day for an arid summer scenario with no vegetation on the ground. The points in the data represent measurements, so that C_n^2 varies widely even at a particular time of data. However, the chart shows the range of C_n^2 values and shows that midday is when the C_n^2 values are at a peak. Dawn and dusk are when the turbulences are the most calm. Finally, turbulence at night varies widely and can be significant.



Figure 6.13 Two sets of bar targets imaged through significant turbulence.

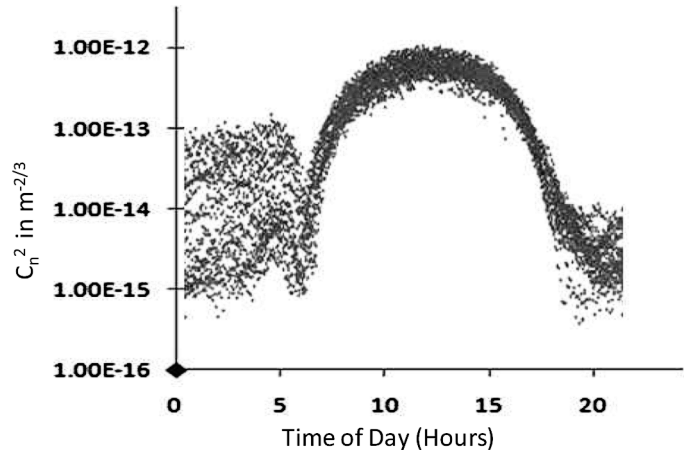


Figure 6.14 C_n^2 as a function of time of day in arid summer. (From [16].)

The values in the Weiss-Wrana data can be used to determine whether a sensor is "turbulence limited" or "optics limited." Figure 6.15 shows the 50% probability of target identification as a function of C_n^2 for a scout infrared sensor. This chart shows both MWIR and LWIR performance for an 8-inch aperture and a 640×480 focal plane with 20-μm detectors. The system is diffraction limited and the ranges shown to the left of the chart correspond to "optics-limited" operation. As the C_n^2 increases, the turbulence blur associated with the short exposure MTF, (6.26), increases and the target identification range decreases. Note that when turbulence is low and the system is optics limited, the MWIR sensor has an identification range much higher than that of the LWIR sensor. The MWIR diffraction spot at 4 μm is much smaller than the LWIR diffraction spot at 10 μm. Eventually, LWIR range performance is higher than MWIR performance because the turbulence affects the LWIR sensor less. This type of turbulence analysis chart appears different for various sensor designs.

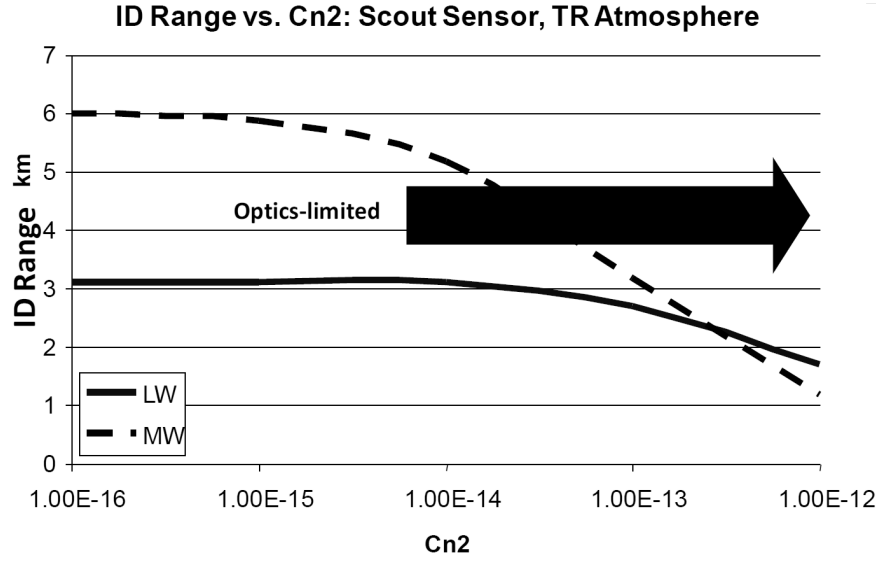


Figure 6.15 Target identification range as a function of C_n^2. (From [17].)

The image quality from wide FOV systems is not typically degraded by turbulence. The resolution of these systems is typically much lower and the turbulence blur is relatively small compared to the sensor blur. Turbulence can have a dramatic effect on narrow FOV systems where the turbulence blur dominates the performance of the system. In recent years, a class of sensors called "ultra-narrow FOV" or UNFOV sensors has been proliferating widely with fields of view of much less than 0.5 deg. These systems have such high resolution and large apertures that they become turbulence limited quite often. For this reason, correction of turbulence effects has been of high interest in recent years for sensor development programs.

6.11 Exercises

6.1 A CO_2 laser (10.6 μm) provides a peak intensity of 10 W/sr and is directed toward a detector 1 km away. Using Figure 6.2, provide a rough estimate of the laser irradiance on the detector on a typical day.
6.2 A particular atmospheric extinction coefficient for a given wavelength is 3.2 per kilometer. Calculate the transmission for a 4-km path length.
6.3 Determine the dominant absorption component of light with (a) a 4.2-μm wavelength, (b) a 6-μm wavelength, and (c) a 9-μm wavelength.
6.4 Determine the most sensitive wavelength of light for a material with a strong bandgap energy of 0.1 eV.
6.5 Describe the type of scattering pattern caused by fog droplets for (a) a He-Ne laser at 0.6328 μm, (b) an Nd:YAG laser at 1.06 μm, and (c) a CO_2 laser at 10.6 μm.
6.6 Calculate the temperature structure parameter for a 10K differential temperature between a 2 m vertical path length.
6.7 Calculate the index structure parameter for a typical desert day using the temperature structure parameter found in Exercise 6.6.
6.8 Determine the index structure parameter for a cool midnight with a temperature of 15°C, a relative humidity of 30%, and calm (negligible) winds.
6.9 Determine the ground-based index structure parameter given that the index structure parameter at a 25 m height above the ground is 4.0×10^{-16} $m^{-2/3}$.
6.10 A longwave infrared system (10-μm average wavelength) is operating in a farfield, long-exposure scenario. The ground-based index structure parameter is estimated to be 1×10^{-15}. The sensor is at an altitude of 25 m and is viewing a target that is 5 km away in a horizontal direction. Determine the long-exposure turbulence MTF of the system. Determine the MTF for EO and MWIR systems under the same conditions.

References

[1] Neilburger, M., J. Edinger, and W. Bonner, *Understanding Our Atmospheric Environment*, W. H. Freeman, San Francisco, CA: pp. 36–41, 1982.

[2] Streetman, B., *Solid State Electronic Devices*, Saddle River, NJ: Prentice-Hall, 1980.

[3] Waldman, G., and J. Wooton, *Electro-Optical Systems Performance Modeling*, Norwood, MA: Artech House, 1993.

[4] Seyrafi, K., and S. Hovanessian, *Introduction to Electro-Optical Imaging and Tracking Systems*, Norwood, MA: Artech House, p. 79, 1993.

[5] Seyrafi, K. and S. Hovanessian, *Introduction to Electro-Optical Imaging and Tracking Systems*, Norwood, MA: Artech House, p. 80, 1993.

[6] Military Handbook, *Quantitative Description of Obscuration Factors for Electro-Optical and Millimeter Wave Systems Metric*, DOD-HDBK-178(ER), pp. 2–11, 1986.

[7] Sadot, D., S. Shamriz, I. Dror, and N. Kopeika, "Prediction of Overall Atmospheric Modulation Transfer Function With Standard Weather Parameters: Comparison With Measurements With Two Imaging Systems," *Optical Engineering*, Vol. 34, No. 11, 1995.

[8] Sadot, D., and N. Kopeika, "Forecasting Optical Turbulence Strength on the Basis of Macroscale Meteorology and Aerosols: Models and Validation," *Optical Engineering*, Vol. 31. No. 2, p. 200, 1992.

[9] Tatarski, V., *Wave Propagation in Turbulent Medium*, Columbus, OH:: McGraw-Hill, 1961.

[10] Goodman, J., *Statistical Optics*, New York City, NY: Wiley, 1985.

[11] Sadot, D., and N. Kopeika, "Imaging Through the Atmosphere: Practical Instrumentation-Based Theory and Verification of Aerosol Modulation Transfer Function," *Journal of the Optical Society of America A*, Vol. 10, No. 1, p. 1017, 1993.

[12] Holst, G., *Electro-Optical Imaging System Performance*, Orlando, FL: JCD Publishing, 1996.

[13] Lenoble, J., *Atmospheric Radiative Transfer*, Hampton, VA: Deepak Publications, 1993.

[14] LOWTRAN Manual (Software Guide), Ontar Corp, Mass, 1995.

[15] Hodgkin, V., B. Kowalewski, S. Moyer, B. Teaney, N. Devitt, J. Fanning, et al., "Comparison of MWIR and LWIR Signatures of Conventional and Human Targets and Rural and Urban Environments," *Military Sensing Symposium*, Orlando, FL, 2006.

[16] Weiss-Wrana, K., "Turbulence Statistics Applied to Calculate Expected Turbulence-Induced Scintillation Effects on Electro-Optical Systems in Different Climate Regions," *Proceedings of SPIE*, Vol. 5237, pp. 1–12, 2004, Bericht FGAN-FOM 2003/11.

[17] Richardson, P., and R. Driggers, "Atmospheric turbulence effects on 3rd generation FLIR performance," *SPIE Proceedings, Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, G. C. Holst (ed), Vol. 6207, 2006.

CHAPTER 7

Optics

The optics of an infrared or EO system collects radiation from a scene and projects an image of the scene onto the system detector array. In the design of the optics, it is important to understand the limitations that optical components introduce to the overall system. Among the parameters defined by the optics are spatial and spectral properties of the sensor system, resolution, and FOV. While the optics play a key role in defining these variables, other sensor system components, such as detector array characteristics, also have a strong system influence.

There are four basic goals in the design of an optical system. The first is to maximize overall system performance. The second is to maximize the resolving power for the desired task while maintaining good area coverage. The third goal is to maximize the amount of image flux that is collected, and the fourth goal is to minimize system complexity and weight. We focus here on the first two goals.

The effect of optics (Figure 7.1) on overall system performance can be described by (1) the diffraction effects associated with the optics, (2) the geometric blur caused by imperfect imaging lenses, and (3) the amount of signal collected by the optical system. Before discussing these effects, we provide background material on the interaction of light and glass.

7.1 Light Representation and the Optical Path Length

In the analysis of optical systems, one must consider how to represent light at each stage of the analytical process. There are three primary methods, each with its own attributes. The first concept of light is that it is made up of a stream of photons. Each photon has energy:

$$E = h\nu \tag{7.1}$$

where E is the energy in joules or electron volts of the photon, h is Planck's constant 6.6252×10^{-34} J-s or equivalently 4.134×10^{-15} eV-s, and ν is the frequency of light in hertz. This representation is convenient in discussions where light interacts with a material such as the detector material.

The second representation of light is that of an electromagnetic wave:

$$E = E_0 \sin(\omega t - k z) \tag{7.2}$$

where E_0 is the wave amplitude, ω is the angular frequency in units of radians per unit of time, t is time, k is the propagation constant $2\pi/\lambda$, z is distance in the direction of propagation, and λ is the wavelength. Here light is referred to as a traveling wave because it is a function of both time and space. The wave nature of light

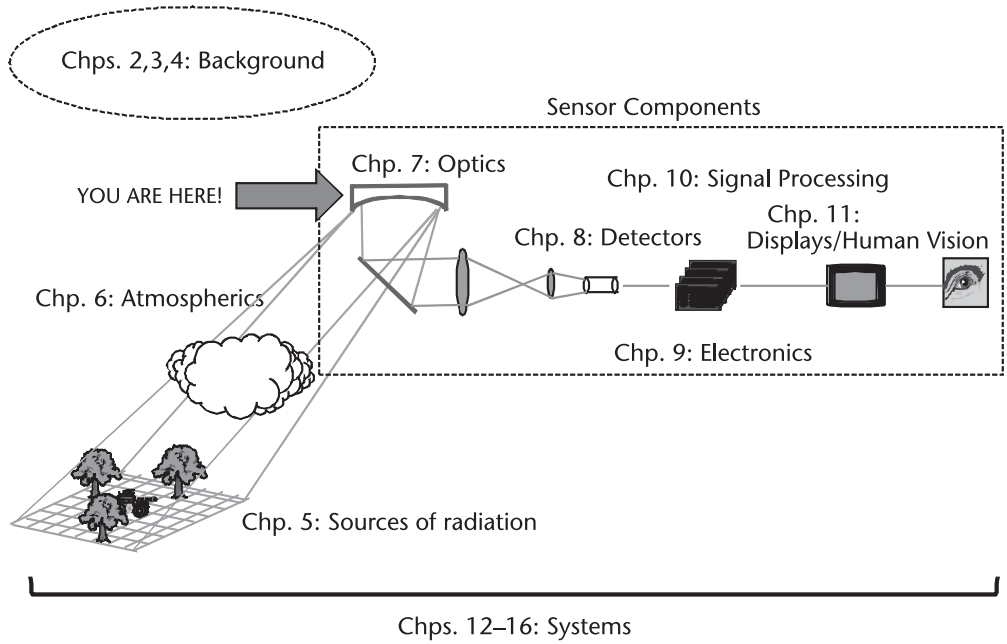


Figure 7.1 Optics in the sensor system.

is very different from the particle or photon nature of light, just as the molecular properties of water are different from the wave characteristics of ocean water at the beach.

The third representation of light is through the use of rays, as shown in Figure 7.2. Rays are defined as lines in the direction of propagation that are everywhere

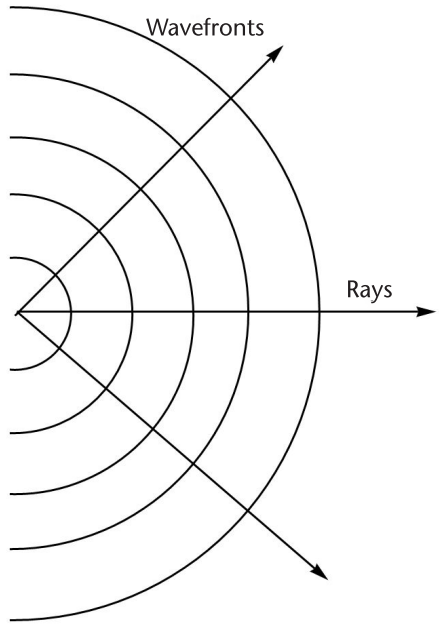


Figure 7.2 In an isotropic medium rays are perpendicular to the wavefronts.

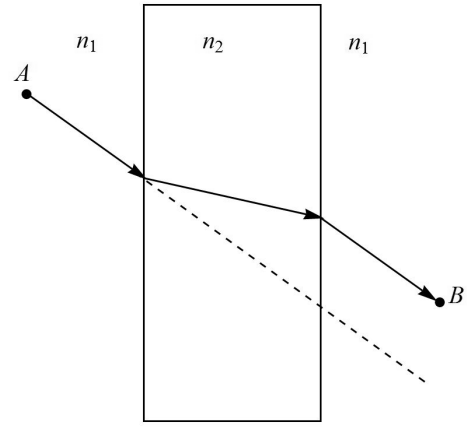


Figure 7.3 Optical path.

perpendicular to the wavefronts, which are defined in Chapter 4 directly under (4.1). Rays depict light as traveling in straight lines and allow optical systems to be analyzed with simple techniques in a field called *geometric optics*. These three treatments of light have been summed up by Saleh and Teich[25]: "The theory of quantum optics provides an explanation of virtually all optical phenomenon. The electromagnetic theory of light provides the complete treatment within the confines of classical optics Ray optics is the limit of wave optics when the wavelength is short."

The path of the ray is not the path of least distance if more than one index of refraction is involved (Figure 7.3). The speed of light in a material is c/n, where c is the speed of light in a vacuum and n is the material's index of refraction (or refractive index). Fermat's principle[22] can be restated in terms of an *optical path length* (OPL):

$$OPL \equiv \int_A^B n(s)\,ds \tag{7.3}$$

The refractive index is given as function of position s along the path. It can be shown that Fermat's principle corresponds to the path with the smallest OPL. The refined version of Fermat's principle states that the actual path taken by the light ray is such as to make its optical path length equal, in the first approximation, to other paths closely adjacent to the actual one.

7.2 Reflection and Snell's Law of Refraction

When a ray of light strikes a boundary of two dissimilar, transparent surfaces, it is divided into two components. One component is transmitted into the second medium and the other is reflected from the boundary surface. A diagram of this is shown in Figure 7.4, where the symbols I, R, and T denote the incident, reflected,

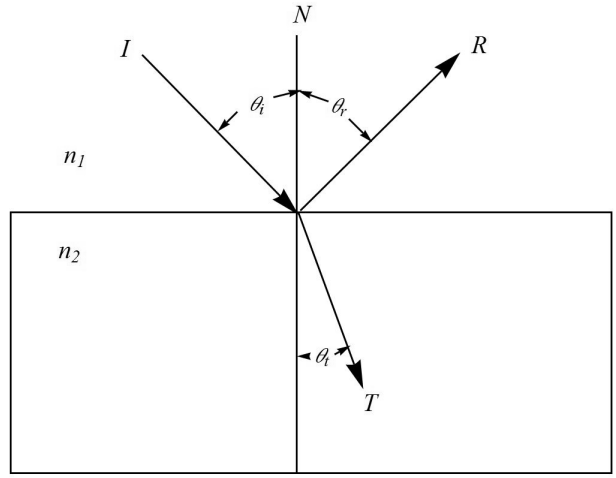


Figure 7.4 Reflection and refraction at an interface.

and transmitted rays, respectively. The normal to the surface is denoted by N and the indices of refraction outside and inside the medium are denoted by n_1 and n_2. The angles of incidence, reflection, and transmission are denoted by θ_i, θ_r and θ_t, respectively. In each case the angle is measured from the normal. The *plane of incidence* is the plane defined by the incident ray and the normal. In the case shown the transmitted ray is bent toward the normal and this indicates that $n_2 > n_1$. The law of reflection[13] states that the reflected ray lies in the plane of incidence, and the angle of reflection equals the angle of incidence:

$$\theta_i = \theta_r \tag{7.4}$$

The law of refraction[13], also known as Snell's law, states that the refracted ray lies in the plane of incidence and the sine for the angle of refraction bears a constant ratio to the sine for the angle of incidence:

$$n_i \sin\theta_i = n_t \sin\theta_t \tag{7.5}$$

where n is the index of refraction. This means that when light passes from one medium to another, the speed of light as well as the propagation direction changes. The index of refraction is defined by the relationship $n_m \equiv c/v_m$ where c is the speed of light in vacuum and v_m is the speed of light in medium m.

Equations (7.4) and (7.5) are fundamental to geometrical optics and may be derived[21] from Fermat's principle or from the Huygens' wavelet approach discussed in Chapter 4. The law of reflection given by (7.4) accurately describes a smooth, polished surface, which is said to be *specular*. *Lambertian surfaces* are rough (see Figure 5.5) and for those surfaces the incident ray is not reflected as shown in Figure 7.4 but instead is scattered through a variety of angles. Optical surfaces are typically specular and (7.4) accurately describes their reflection properties.

The index of refraction of a vacuum is 1 and ordinary air has an index of refraction of approximately 1.0003. Refractive indices given in glass catalogs are usually with respect to air, so vacuum calculations in lens design codes use 1/1.0003 as

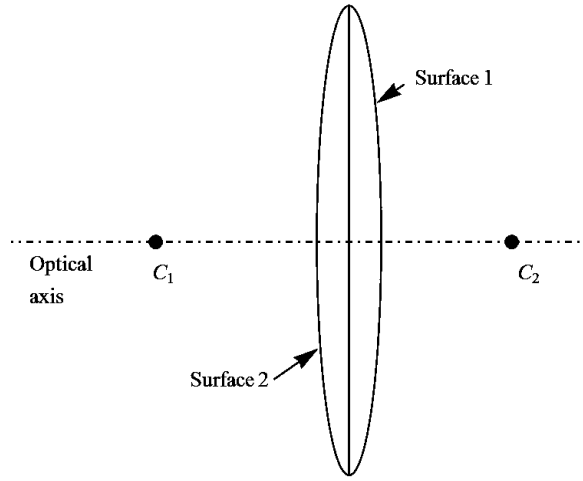


Figure 7.5 The optical axis.

the refractive index of the vacuum. The index of refraction of a material is also a function of wavelength, but correction can be applied in the design of an imaging system to minimize spectral variation over a wavelength band of interest. Typical window glass has a refractive index of 1.5, whereas water has a refractive index of around 1.33.

7.3 The Thin Lens Ray-Tracing Rules and Gauss's Equation

Figure 7.5 illustrates the concept of an *optical axis*. A simple lens typically has two spherical surfaces. Each sphere has a center labeled C_1 and C_2 as in the figure. For a simple thin lens, the optical axis is the line connecting C_1 and C_2. A high-quality lens often consists of multiple thin lenses. In that case, proper lens assembly requires the optical axis of each individual lens to be coincident with a single line that is the optical axis for the system of lenses. Modern optical techniques allow for nonspherical surfaces. However, each of the lens surfaces can be generated by rotating some curve about an axis that is taken to be the optical axis. In constructing a lens with nonspherical surfaces, care is used to ensure that the optical axis for each surface is coincident.

The top part of Figure 7.6 illustrates several types of convex lenses and a method for representing convex lenses in optical diagrams. *Convex lenses* are lenses that get thicker closer to the optical axis and are thickest at the optical axis. The bottom left side of Figure 7.6 shows that collimated light incident from the left and parallel to the optical axis is brought to a focus on the optical axis, which is called the *image space focal point*. The bottom right side of the figure shows that a point source located at the object space focal point produces a collimated beam exiting the lens. *Collimated light* is light in which the light rays are parallel to each other. Because rays are perpendicular to the wavefront, collimated light corresponds to a plane wave.

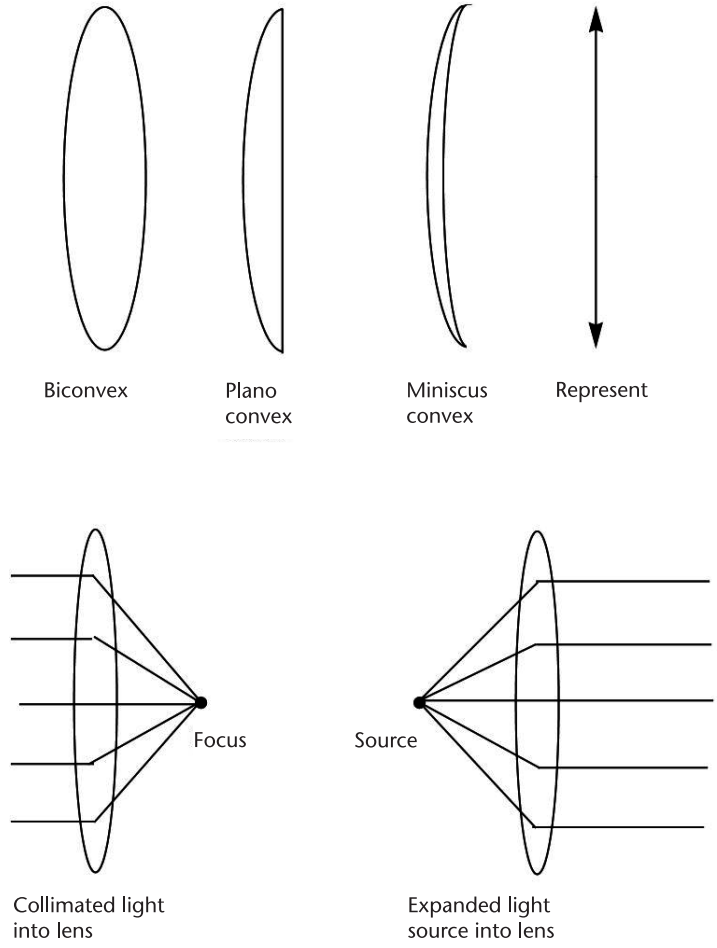


Figure 7.6 Convex lens nomenclature, representation, and behavior. Collimated light parallel to the optical axis incident on a convex lens is brought to a focus on the optical axis. A convex lens produces collimated light from a source at a focal point.

The top part of Figure 7.7 illustrates several types of concave lenses and a method for representing concave lenses in optical diagrams. *Concave lenses* are lenses that get thinner closer to the optical axis and are thinnest at the optical axis. The bottom left side of the figure shows that collimated light incident from the left and parallel to the optical axis appears to diverge from a point on the optical axis, which is called the *object space focal point*. The bottom right side of the figure shows that a source that is converging on the *image space focal point* results in a collimated beam exiting the lens.

Figure 7.8 illustrates the basic concept of image formation. An object can be thought of as a collection of sources that emit or reflect light in many directions. As illustrated in the figure, the convex wavefront that emanates from the object point A on the optical axis is brought to a focus A' on the optical axis. The action of a convex lens is to take diverging rays and bend them back toward the optical axis. The convex lens has a limited ability to do this, characterized by the focal length of the lens. In Figure 7.8 the object is far enough away so the lens has enough power

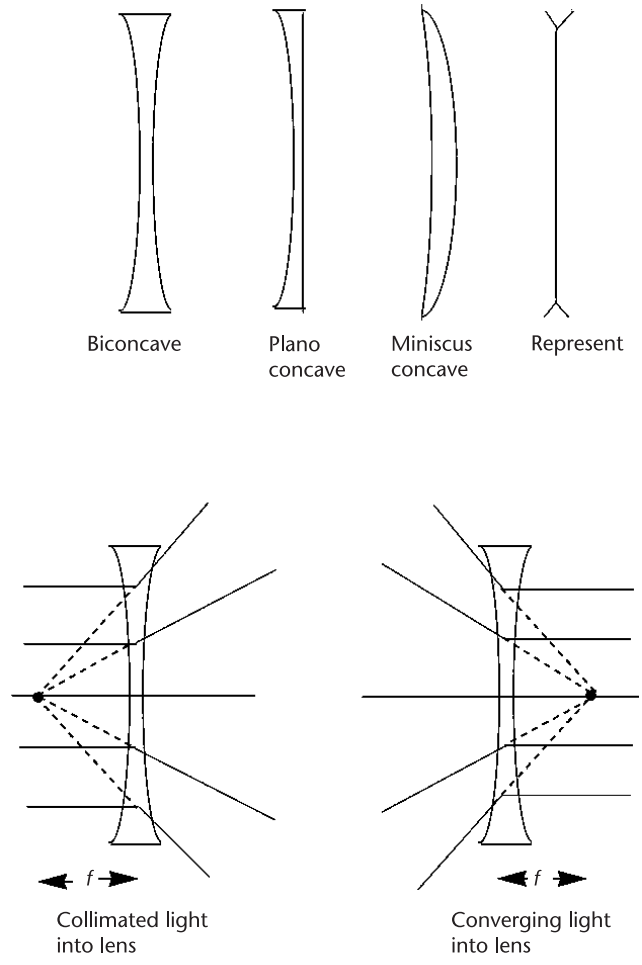


Figure 7.7 Concave lens nomenclature, representation, and behavior. Collimated light parallel to the optical axis incident on a concave lens appears to diverge from the image space focal point. Light converging to the object space focal point is converted to collimated light parallel to the optical axis.

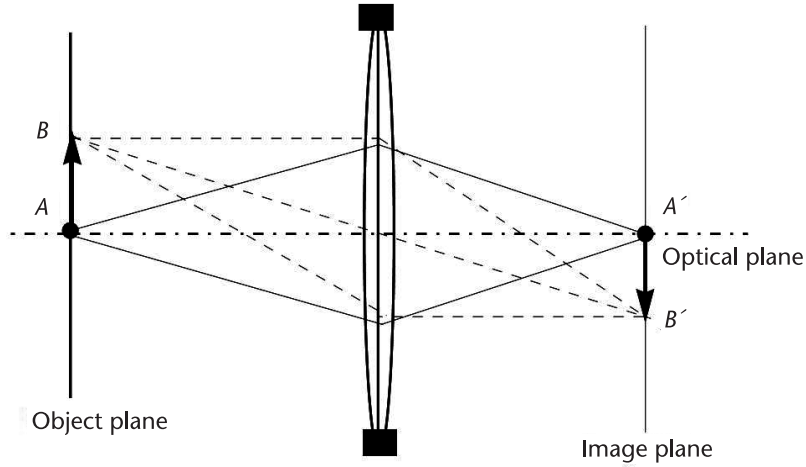


Figure 7.8 Image formation by a convex lens.

to bend the rays emerging from point A on the optical axis back to point A' on the optical axis. The plane perpendicular to the optical axis at point A' is called the *image plane* or *focal plane*, and the plane perpendicular to the optical axis at the point A is called the *object plane*. The off-axis point B is imaged to the point B' on the image plane.

The ray traces shown in the lower part of Figure 7.8 take place in a *meridional plane,* also called the *tangential plane*. This is the plane defined by an off-axis point, for example, B, and the optical axis. Typically, ray diagrams are done in the meridional plane. The lens surface depicted in Figure 7.8 is the intersection of the lens surface with the meridional plane. The *sagittal plane* is perpendicular to the meridional plane at the optical axis.

Figure 7.9 illustrates a model for tracing rays through a thin convex lens. Focal points f and f' are on the optical axis. The lens is assumed to be thin so that all of the bending produced by the lens is assumed to take place on a single plane, called the *principal plane*, that is perpendicular to the optical axis. In reality, the radiation is bent at each surface of the lens but for many purposes assuming all bending takes place at the principal plane is adequate for a first-order understanding of the optical system. Let O denote the point where the principal plane intersects the optical axis. Then if the index of refraction is the same on both sides of the lens, the unsigned distances $f\,O$ and $O\,f'$ are equal.[13] A first-order model for a thin convex lens is summarized by three rules: (1) The ray that passes through the center of the lens passes through undeviated, (2) the ray incident on the lens that is parallel to the optical axis passes through the focal point f', and (3) a ray that passes through the focal point f emerges from the lens traveling parallel to the optical axis.

The first rule is readily understood by examining Figure 7.3. Using Snell's law, the direction of the ray emerging from a plane parallel plate is parallel to but displaced from the incident ray. As the refracting sheet gets thinner, the displacement gets smaller. Examination of Figure 7.9 shows that near the optical axis, a thin lens is approximately two parallel planes. For some calculations the displacement observed in Figure 7.3 is negligible. Another way to see that the ray through the center of the lens goes through undeviated is to realize that ray 2 in Figure 7.9 is bent

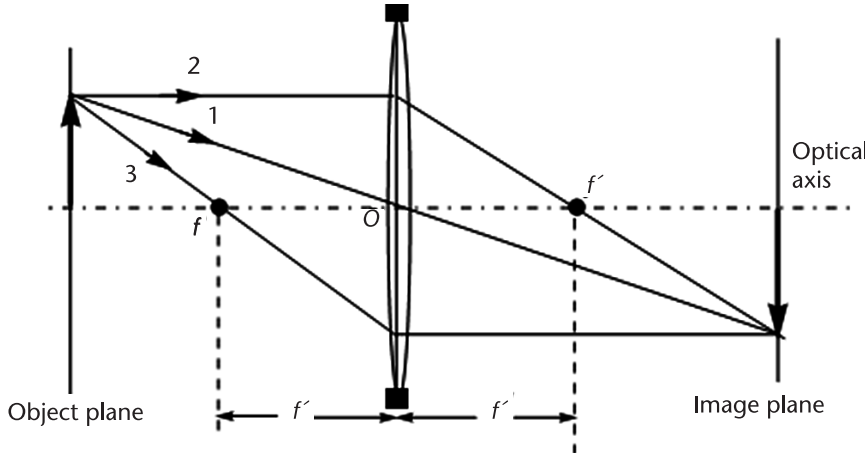


Figure 7.9 Ray-tracing rules for a thin convex lens.

down by the lens and ray 3 is bent up by the lens. Since the direction of bending must be a continuous function of where the ray goes through the lens, somewhere a ray passes through the lens undeviated. By symmetry, that point is the center of the lens. The second and third rules can be thought of as approximate experimental results obtained with laser beams. These rules are expressed in the bottom portion of Figure 7.6.

Figure 7.10 illustrates a model for tracing rays through a thin concave lens. Focal points f and f' are on the optical axis. In Figures 7.9 and 7.10 the focal point labeled f has these properties: (1) A ray incident on f leaves the lens in a direction parallel to the optical axis and (2) a ray back traced through f' exits the lens in a direction parallel to the optical axis. The lens is assumed to be thin so that all of the bending produced by the lens is assumed to take place on a single plane called the *principal plane*, which is perpendicular to the optical axis. Let O denote the point where the principal plane intersects the optical axis. If the index of refraction is the same on both sides of the lens, then the unsigned distances Of' and Of are equal.[13] A first-order model for a thin concave lens is summarized by three rules: (1) The wray that passes through the center of lens passes through undeviated, (2) the ray incident on the lens that is parallel to the optical axis passes through the focal point f', and (3) a ray pointed toward the focal point f emerges from the lens traveling parallel to the optical axis. Arguments similar to those already given for a convex lens justify the ray-tracing rules for a thin concave lens.

Figure 7.11 illustrates the *focal plane* concept and how a distant point source is imaged on the focal plane. The collimated rays in the top part of the diagram come from a distant point on the optical axis. The collimated rays in the lower part of the diagram come from a distant point source that subtends an angle θ with respect to the optical axis. In reality, the surface on which a distant off-axis point is imaged is curved and the tangent of that surface is perpendicular to the optical axis. For many purposes it is a good approximation to replace the curved surface with the plane surface that is perpendicular to the optical axis and that is the model usually used to predict the location of the image using first-order optics. Examining the lower

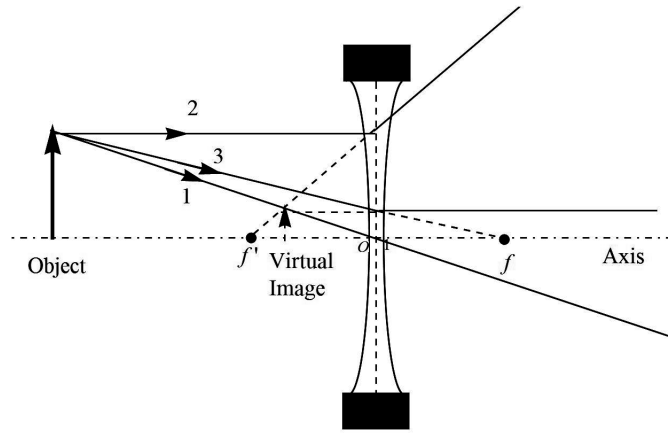


Figure 7.10 Ray-tracing rules for a thin concave lens.

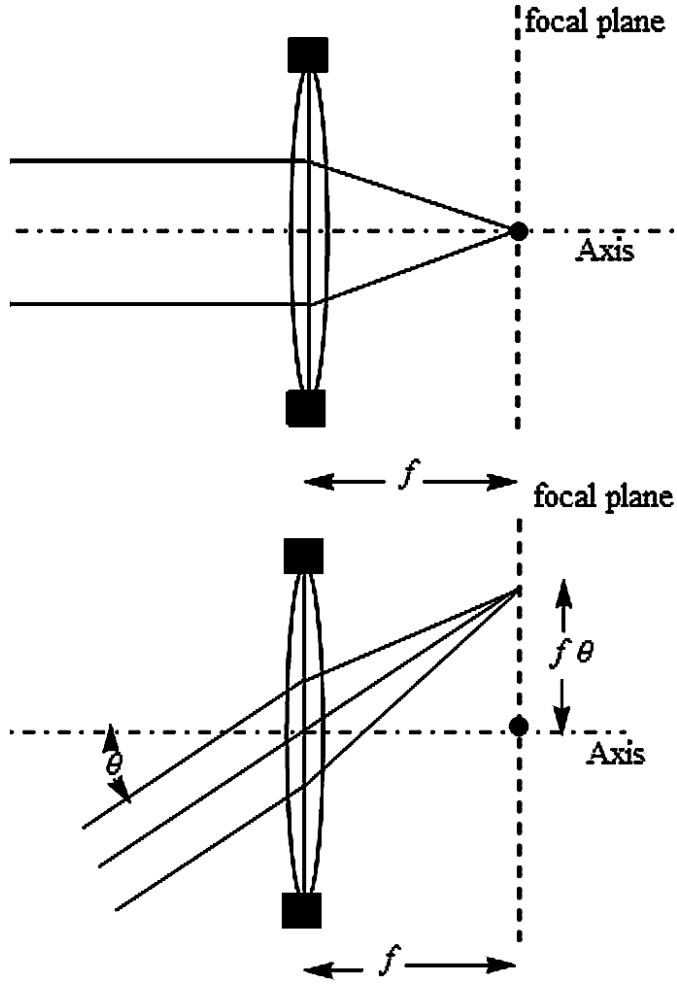


Figure 7.11 Imaging of a point source on and off axis.

part of the figure, the imaged point is a distance $f\,tan\theta$ away from the optical axis. Typically θ is small so $tan\theta \approx \theta$ and so, to a good approximation, the distant point object is imaged a distance $f\theta$ from the optical axis.

Snell's law can be used to calculate how far away the focal points are from the principal plane of a thin lens[21]:

$$\frac{1}{f} = \frac{(n_l - n_m)}{n_m}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{7.6a}$$

where n_l and n_m are the refractive indices of the lens material and of the medium, and $1/R_1$ and $1/R_2$ are the curvatures of the two surfaces. The indices 1 and 2 correspond, respectively, to the first and second refractive surfaces encountered by the radiation. If the center of curvature is to the right, then the curvature is regarded as positive; otherwise, it is regarded as negative. Typically, the medium on both sides of a lens is air so $n_m = 1$ and n_l is the index of refraction with respect to air, which

is the way indices of refraction are normally given in glass catalogs. With $n_m = 1$, (7.6a) becomes

$$\frac{1}{f} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{7.6b}$$

which agrees with a form of the equation given by Hecht and Zajac[10]. This equation can be generalized to accommodate a thick lens in air[8] (http://en.wikipedia.org/wiki/Lens_(optics))

$$\frac{1}{f} = (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2} + \frac{(n_1 - 1)d}{n_l R_1 R_2}\right) \tag{7.6c}$$

where d is the thickness of the lens (the distance along the axis between the two surface vertices).

Equations (7.6a), (7.6b), and (7.6c) allow one to calculate the focal length of a lens and they are collectively referred to as the *lensmaker's equation*. The interesting thing to note is that for a given value of n_l and n_m there is considerable freedom in the choice of R_1,R_2, and d. This freedom can be used to minimize aberrations.

Given the locations of the principal plane and the focal points, Figure 7.12 shows how to find the exit ray for an arbitrary ray incident on a thin convex lens. The top part of the figure shows an arbitrary ray A incident on a convex lens. As shown in the bottom part of the figure, the ray A is traced backward to point B. Using our ray-tracing rules, rays 1, 2, and 3 are drawn from B and are imaged at

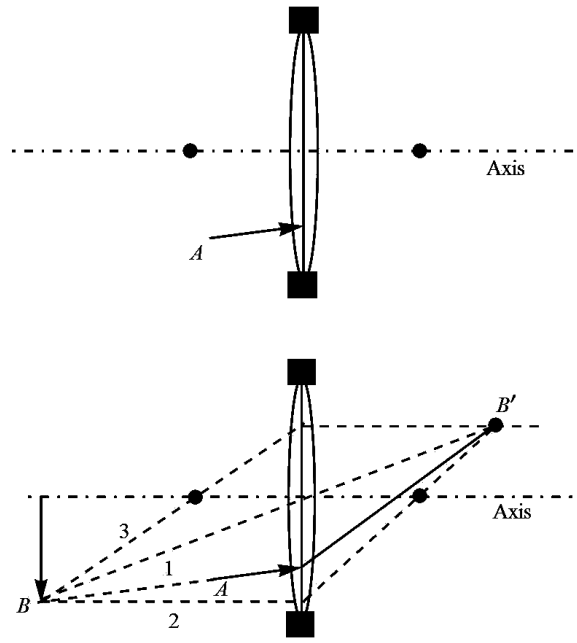


Figure 7.12 Ray tracing of an arbitrary ray through a thin convex lens.

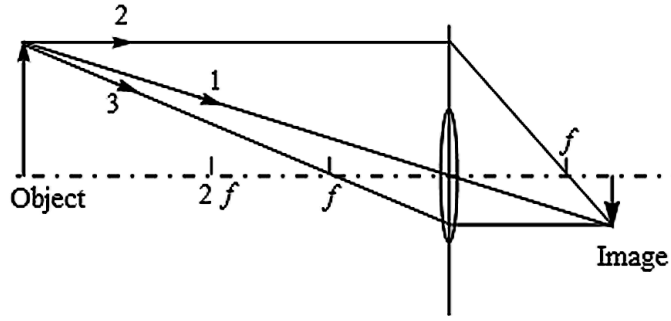


Figure 7.13 The principal plane extends far beyond the physical limits of the lens.

point B'. All the rays emanating from B are imaged at point B', which implies that ray A goes to point B' when it exits the thin lens.

A similar procedure can be used to find the direction an arbitrary ray A exits when incident on a concave lens. One simply back traces the incident ray A to a convenient point B as was done in Figure 7.12. Then use rays 1, 2, and 3 as illustrated in Figure 7.10 to find the virtual image B' of B. Similar procedures can be used to trace an arbitrary ray from a convex mirror or a concave mirror.

If the object is larger than the lens, then the ray parallel to the optical axis (ray 2 in Figure 7.13) will miss the lens. However, the image position could still be determined by using rays 1 and 3 in that figure. As shown in Figure 7.13, according to the thin-lens ray-tracing model, the principal plane can be extended far beyond the physical limits of the lens and this allows us to use ray 2 as well.

Figure 7.14 shows how a thin convex lens behaves as the distance between the object and the principal plane varies from infinity to less than one focal length. In Figure 7.14(a) we have a point object on the optical axis at infinity that is brought to a focus at the focal point. As the object moves from infinity to $2f$, as illustrated in Figures 7.14(b) and (c), the inverted real image gets larger and moves from f to $2f$. When the object is at $2f$ the inverted image is the same size as the object. Figure 7.14(d) shows what happens when the object is between $2f$ and f away from the principal plane. In this case we have a real inverted image that is larger than the object. Figure 7.14(e) shows that when the object is a distance f from the lens, the rays from the object are diverging so much that the lens can no longer bring the rays into focus and there is no image. If the object is brought closer to the lens than one focal length, as illustrated in Figure 7.14(f), then the rays exiting the lens diverge and there is no real image. However, if a person looks through the lens, he or she will see a virtual image of the object that is erect and is larger than the object.

An image can be *real* or *virtual*. Examining Figure 7.14(a–d), the rays from a point on the object converge to a point and we have a *real image*. If film or a detector array were placed such that it included the point where the rays converge and were perpendicular to the optical axis, then an image would be recorded. In Figure 7.14(e) the exiting rays are parallel and there is no virtual or real image. In Figure 7.14(f), the rays exiting the lens diverge but appear to come from a point on a *virtual image*. An image-forming optical system[22] gathers a beam of light diverging from an object point and transforms it into a beam that converges toward or

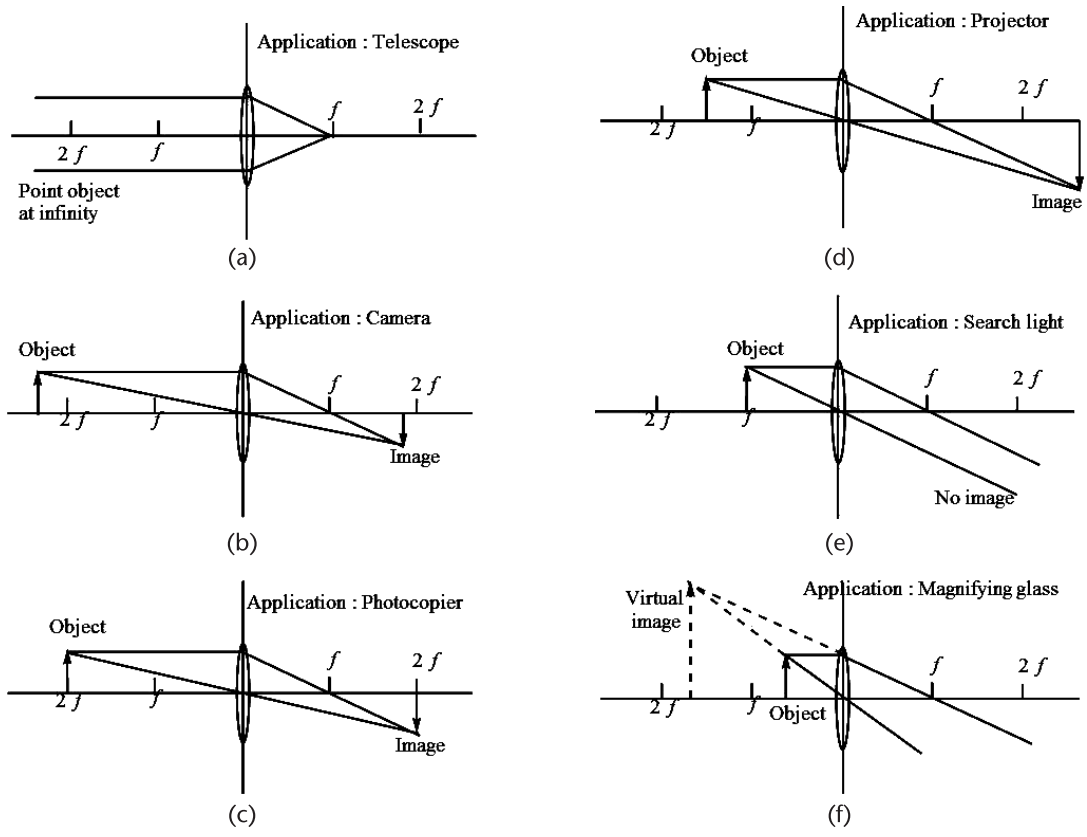


Figure 7.14 Summary of simple convex lens behavior.

diverges from another point, thus producing an image. If the beam converges to a point, a *real image* is produced. If the beam diverges from a virtual point, as in Figure 7.14(f), a *virtual image* is produced.

In Figure 7.15 the distance between the object and a concave lens varies from beyond $2f$ to a distance less than f. Observe that the image is always virtual and is always smaller than the object. The virtual image size gets larger as the object moves closer to the lens. This is easily understood by realizing that, given the object is to the left of the lens, the backward projection of ray 2 is independent of the object's position but the angle of ray 1 gets steeper as the object gets closer to the lens. In the limit as the object approaches the principal plane, the object and the virtual object coincide.

Given the focal length and object location, using the ray-tracing rules we can determine the image location. Gauss's law is an algebraic relationship between the object distance, focal length, and image distance and it can be derived from the ray-tracing rules already developed.

Examining Figure 7.16, we get

$$tan\,\theta = \frac{h_o}{s_o} = \frac{h_i}{s_i} \;\Rightarrow\; \frac{h_i}{h_o} = \frac{s_i}{s_o} \tag{7.7a}$$

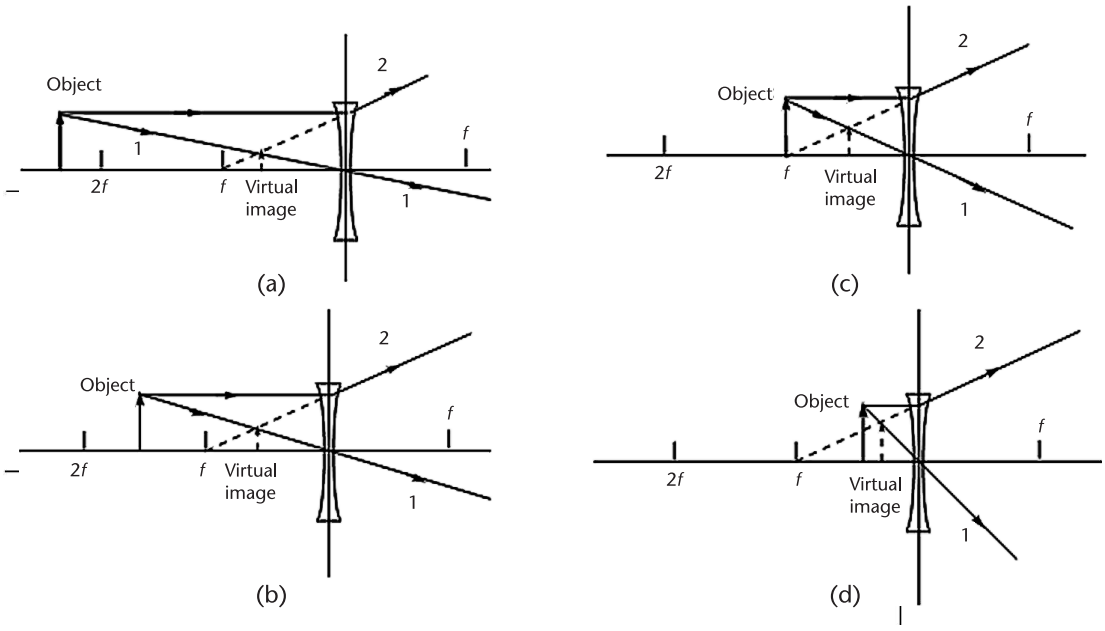


Figure 7.15 Summary of concave lens behavior.

$$tan\,\alpha = \frac{h_o}{s_o - f} = \frac{h_i}{f} \;\Rightarrow\; \frac{h_i}{h_o} = \frac{f}{s_o - f} \tag{7.7b}$$

Comparing (7.7a) and (7.7b),

$$\frac{s_i}{s_o} = \frac{f}{s_o - f} \tag{7.7c}$$

Equation (7.7c) is a relationship between the distances s_o, s_i and the focal length f of a thin convex lens. It is an exercise in elementary algebra to show that (7.7c) can be written

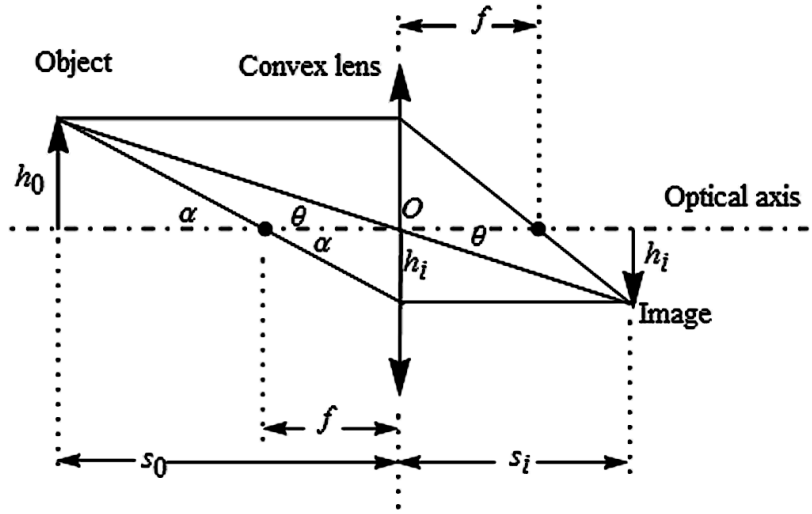


Figure 7.16 Ray trace for deriving Gauss's thin lens equation.

$$\frac{1}{s_o}+\frac{1}{s_i}=\frac{1}{f} \tag{7.7d}$$

Equation (7.7d) is called Gauss's thin lens equation. In Figure 7.16 and (7.7d), it was assumed that s_o, s_i, and f are positive. Some authors use the convention that distances to the left of the principal plane are negative.[26] With this convention (7.7d) becomes

$$-\frac{1}{s_o}+\frac{1}{s_i}=\frac{1}{f} \quad \Rightarrow \quad \frac{1}{s_o}+\frac{1}{f}=\frac{1}{s_i} \tag{7.7e}$$

An interesting application of (7.7e) is to estimate where the image is when s_o is very small and a little to the left of the principal plane. In the notation of (7.7e), s_o is small and negative, which implies that $1/f$ is negligible in comparison. Then (7.7e) implies that s_i is small and negative. So if one has an object at the principal plane, then the image of that object is also at the principal plane (s_o and s_i are both small). The same result is obtained by examination of Figure 7.14. When the object is less than but near f the virtual image is large and far away because then the rays exiting the lens are barely diverging. When the object is between f and the principal plane but near the principal plane, then the image is near the principal plane because now the rays exiting the lens are highly divergent.

We define lateral magnification M_T as the absolute value of the image size divided by the absolute value of the object size, which implies the lateral magnification is always positive. Some authors[8] define a signed magnification where a negative sign corresponds to an inverted image. We prefer to determine whether the image is inverted or right side up from the ray trace. Observe from Figure 7.16 that

$$tan\,\theta=\frac{h_o}{s_o}=\frac{h_i}{s_i} \quad \Rightarrow \quad M_T\equiv\frac{h_i}{h_o}=\frac{s_i}{s_o} \tag{7.8}$$

In Figure 7.16, all quantities are positive so there is no need for absolute value signs in (7.8).

The ray-tracing rules provide a good understanding of what is happening in an optical system. Gauss's thin lens equation allows a more accurate prediction of either s_o, s_i, or f when two of the three quantities are known.

7.4 Spherical Mirrors

The material of a lens determines the wavelength band over which the lens transmits radiation. Typically mirrors reflect over a wider wavelength band than lenses and obey (7.4) over that wavelength band. Also, the index of refraction of lens material depends on wavelength and this implies that the focal length, determined by Snell's law of (7.6), is wavelength dependent so each wavelength has its own focal plane. These are some of the positive attributes of mirrors that are used extensively in optical systems. To understand optical systems that use mirrors, it is useful to know how the ray-tracing rules developed for lenses apply to mirrors.

In Figure 7.17, C denotes the center of curvature of the concave mirror. The optical axis of the spherical mirror passes through C and is perpendicular to the mirror at vertex V. The plane perpendicular to the optical axis at point V is the *principal plane* of the mirror. In our treatment of thin lenses, we made the approximation that all bending took place at the principal plane. Here too we make the approximation that all reflections take place at the principal plane, for example, the reflection of ray 2 is modeled as taking place at the principal plane rather than at the surface Σ where the reflection actually takes place. After reflection from the mirror, ray 2, which is parallel to the optical axis, intersects the optical axis. In the limit, as ray 2 approaches the optical axis, the location where ray 2 intersects the optical axis is called the *focal point f*. Using the law of reflection, it is easy[10] to show that the distance $V\,C$ is twice the distance $V\,F$. The distance $V\,F$ is called the *focal length* of the mirror.

As illustrated in Figure 7.17, the ray-tracing rules for a concave mirror are as follows: Ray 1 from P passing through C gets reflected back on itself; ray 2, parallel to the optical axis, gets reflected through the focal point f; ray 3, passing through the focal point after being reflected at the principal plane, travels in a direction parallel to the optical axis. These three rays enable the image location and size to be found for any object position.

Figure 7.18 illustrates ray-tracing rules for a convex mirror. In the figure the *principal plane* is perpendicular to the optical axis at point V and C is the center of curvature of the convex mirror. Focal point f is the intersection of the optical axis with an extension of the reflection of ray 2 with the optical axis. Here it is assumed that ray 2 at point P is close to the optical axis. The location of image point P' is obtained from rays 1, 2, and 3. Ray 1 is in the direction of C and is reflected back on itself from the principal plane. Ray 2 is parallel to the optical axis and its reflection from the principal plane appears to come from point f. Ray 3 is incident on point f and its reflection is parallel to the optical axis. None of the reflected rays 1, 2, or 3 intersect each other, indicating no real image is formed by a convex lens. Observe that the back-traced reflected rays 1, 2, and 3 intersect at P', which is the location of the virtual image of point P.

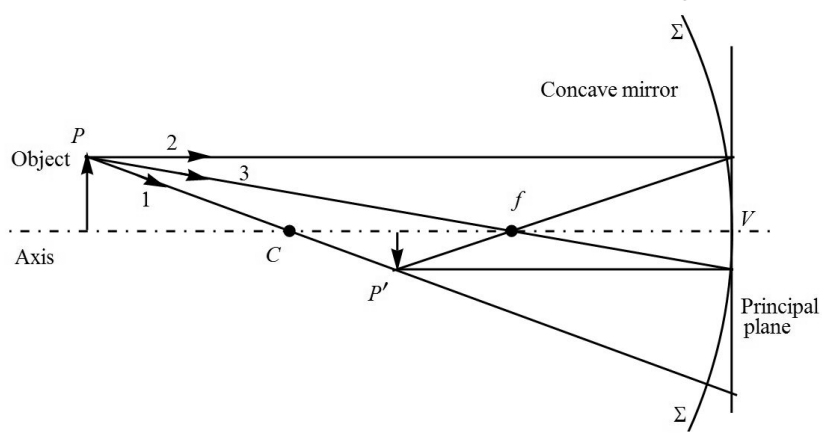


Figure 7.17 Ray-tracing technique for a spherical concave mirror.

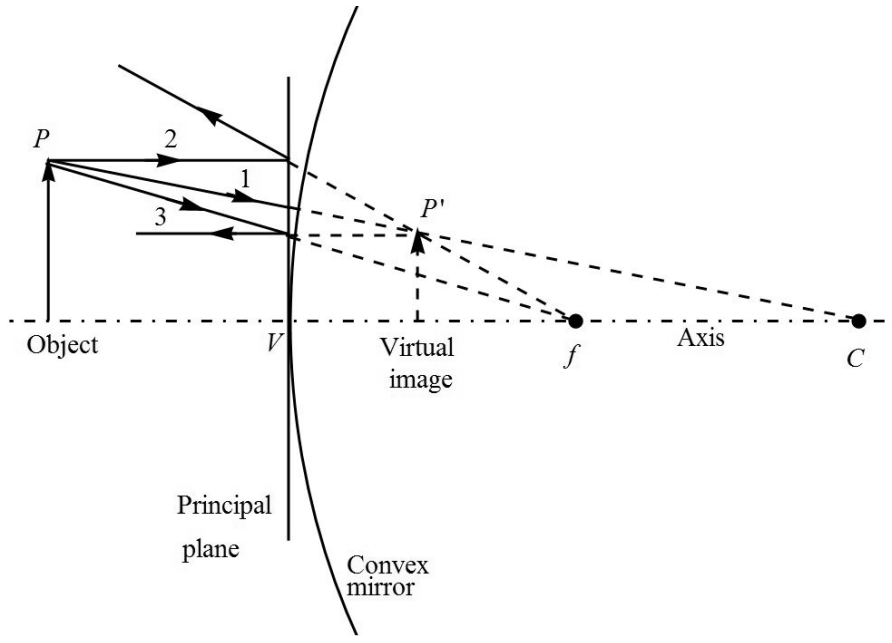


Figure 7.18 Ray-tracing technique for a spherical convex mirror.

7.5 Modeling the Thick Lens

The *first focal point* and *first principal plane* of a convex lens are illustrated in Figure 7.19(a). The optical axis of the lens is conceptually determined using Figure 7.5. Move a point source along the optical axis until the emerging rays are all parallel to the optical axis. When this happens, the location of the point source coincides with the first focal point *f*. Forward trace a ray from the first focal point and back trace the corresponding exit ray. Because refraction is confined to the plane of incidence, these two rays meet at a point. Forward trace all the rays from the first focal point that make it into the lens, and back trace the collimated rays exiting the lens and determine where the rays meet. The points where the rays meet define the *first principal surface,* which is a curved surface that intersects the optical axis at point *H*, called the *first principal point* of the lens. The tangent plane to the first principal surface at the first principal point is perpendicular to the optical axis and is called the *first principal plane*. The distance from the first focal point to the first principal plane is called the *first principal focal length* and is denoted by the symbol *f* in the figure.

The *second focal point* and *second principal plane* of a convex lens is illustrated in Figure 7.19(b). A collimated bundle of rays parallel to the optical axis is incident on the lens from the left. These rays emerge from the lens and converge at the point labeled *f′* called *the second focal point*. Because refracted rays are confined to the plane of incidence, each forward-traced ray intersects each back-traced ray at a point. The collection of all such points defines the *second principal surface*, which in general is curved and intersects the optical axis at point *H′*, the *second principal point*. The tangent plane to the second principal surface at the second principal

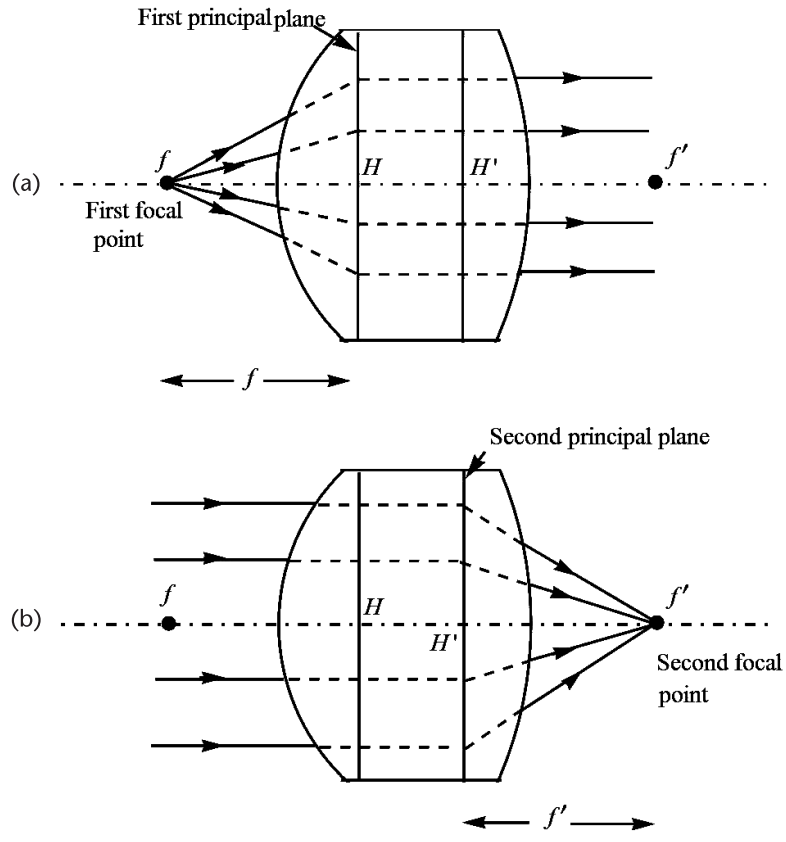


Figure 7.19 Principal planes and focal lengths for a thick lens.

point intersects the optical axis in the *second principal plane*. The second principal plane has the property that it is perpendicular to the optical axis. The distance from the second focal point to the second principal plane is call the *second principal focal length* and is denoted by the symbol f' in the figure.

If the medium to the left of the lens is the same as the medium to the right of the lens, then the two focal lengths are equal[13]:

$$f = f' \tag{7.9a}$$

If the material to the left of the lens is characterized by index of refraction n and the material to the right of the lens is characterized by index of refraction n', then (7.9a) needs modification[13]:

$$\frac{f}{f'} = \frac{n}{n'} \tag{7.9b}$$

and we find that the ratio of the first and second focal lengths is proportional to the ratio of the refraction indices of the medium on the left and right sides of the lens.

The ray that passed undeviated through the center of a thin lens found considerable use in ray tracing. For that reason we seek a corresponding ray for a thick lens. In Figure 7.20a, ray A from a distant off-axis point has the property that the

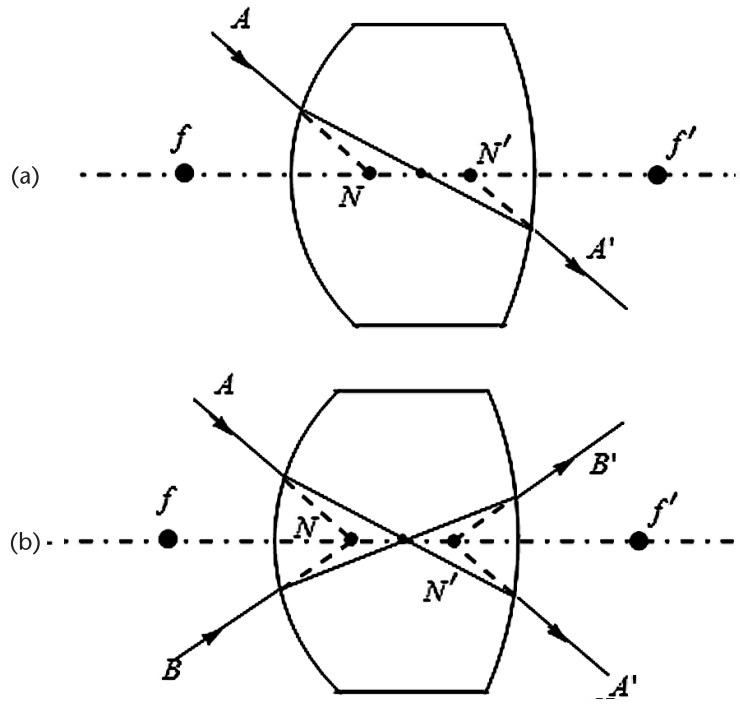


Figure 7.20 Nodal points of a thick lens.

emerging ray is parallel to the incident ray. Incident rays parallel to and above ray *A* are bent down by the lens and the corresponding rays beneath ray *A* would be bent up. Since the amount of bending is a continuous function of how far away the incident ray is from the optical axis at the point where it first encounters the lens surface, there is some ray that will emerge from the lens undeviated and that ray is labeled *A* in the figure. The forward trace of ray *A* intersects the optical axis at *N*. The back trace of ray *A′* intersects the optical axis at the point designated *N′*. The points *N* and *N′* are called the nodal points of the lens.

Figure 7.20b illustrates an important property of the nodal points. Ray *B* is an arbitrary ray heading to point *N*. Then the ray *B′* that emerges from the lens will be parallel to ray *B* and will appear to come from point *N′*. The nodal points are a generalization of the principal point of a thin lens. Any ray incident on the nodal point *N* emerges parallel to the incident ray and appears to come from point *N′*.

If the refractive index to the left of the lens is the same as the refractive index to the right of the lens, then the nodal points have the following important property. The nodal point *N* and the first principal point *H* are the same point. Similarly, the nodal point *N′* and the second principal point *H′* are the same.[13]

Figure 7.21 shows rays *A*,*B*, and *C* incident on a point *P* on the first principal plane. All three rays are then transferred horizontally to the point *P′* on the second principal plane. The incident ray *A* leaves the optical system as ray *A′* traveling parallel to the optical axis. The incident ray *B* leaves the optical system as the ray *B′* that passes through the focal point *f′*. The incident ray *C* that is between rays *A* and *B* leaves the optical system as ray *C′* in a direction between rays *A′* and *B′*.

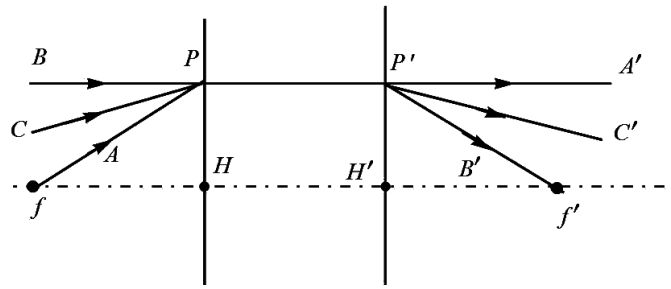


Figure 7.21 Rays incident on the first principal plane are imaged on the second principal plane with unity magnification.

An optical system has[10] six *cardinal points*: f, f', H, H', N, and N'. As shown in Figure 7.22, specification of these six points facilitates first-order ray traces. The medium on the left and right sides of the optical system are assumed to be the same, so the distance from f to H is the same as the distance from H' to f'. Also, H and H' are coincident with the nodal points N and N'. Ray 1 from an off-axis point on the object is incident on point N and emerges in the same direction from point N'. Ray 2 is incident on the first principal plane at point P, gets transferred to point P', and then passes through focal point f'. Ray 3 passes through focal point f, intersects the first principal plane at point Q, gets transferred to point Q', and emerges from the optical system traveling parallel to the optical axis.

7.6 Vergence

Ray tracing and Gauss's equation are two ways to understand the properties of optical systems. A third technique for understanding optical systems makes use of the vergence concept. Utilizing the convention that light travels from left to right, Figure 7.23 illustrates three cases. Figure 7.23(a) shows an expanding beam that is diverging from a point source. Figure 7.23(b) corresponds to *collimated* radiation that is neither converging or diverging, whereas Figure 7.23(c) corresponds to radiation that is coming to a focus. All of these cases can be described with the analytical technique known as[17] *vergence*.

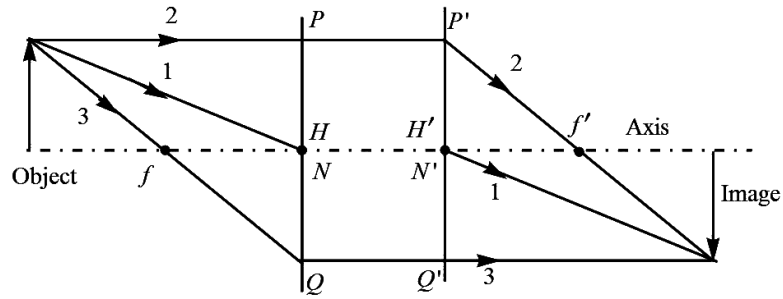


Figure 7.22 Ray tracing through a thick lens.

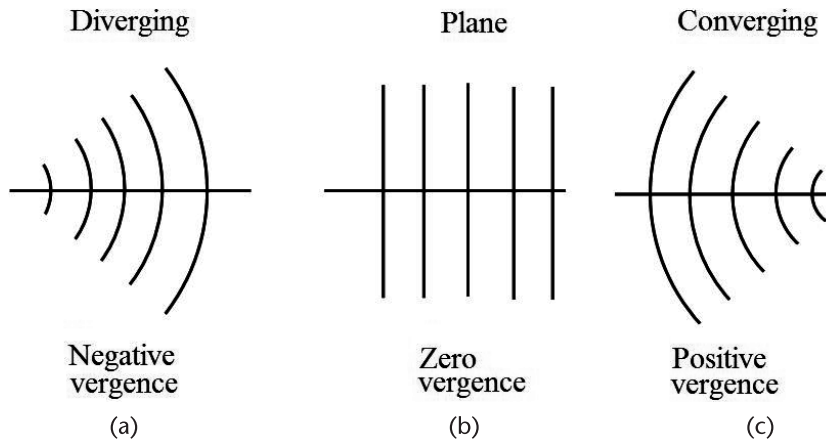


Figure 7.23 Examples of vergence.

The *curvature* C of a wavefront is defined as being the reciprocal of the radius of the wavefront:

$$C \equiv \frac{1}{R} \tag{7.10a}$$

A wave where the center of curvature is to the left as in Figure 7.23(a) has a negative curvature, and a wave where the center of curvature is to the right of the wavefront has a positive curvature.

Vergence can be thought of as wavefront curvature, with a modification for the medium refractive index:

$$V \equiv \frac{n}{R} \tag{7.10b}$$

The units of vergence are 1/meters or *diopters*. Simple rules govern the vergence sign:

1. Light travels from left to right.
2. All distances are measured from a reference surface (where vergence is calculated).
3. Distances to the left of the reference surface are negative; distances to the right are positive.
4. The radius of curvature is measured from a wavefront to the center of curvature: (a) Centers of curvature to the left are negative', (b) centers of curvature to the right are positive', and (c) plane waves have zero vergence.

The sign conventions used here are not universal. For example, [13] considers the vergence in Figure 7.23(a) to be positive rather than negative.

Example 7.1:
Consider the point source shown in Figure 7.24. Determine the vergence of light 2 m from a point source. The reference is the location at which vergence is calculated

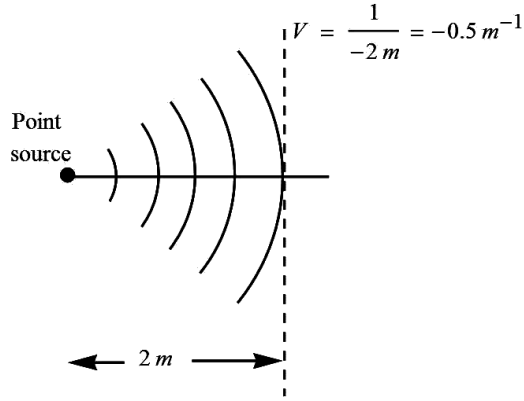


Figure 7.24 Vergence example.

(shown by the dotted line). The center of curvature is to the left of the wavefront, so it is negative and the vergence equals –0.5 m^{-1}.

Example 7.2:
Figure 7.25 illustrates the positive vergence that results in a focus. Determine the vergence 8 cm before the light comes to a focus in air. The vergence is high for high curvature and gets higher as the reference position nears the point source or focus position. Vergence is a function of location and is not a parameter of the system. Here we find the vergence at the reference surface equals 12.5 m^{-1}.

To determine the change in vergence from one plane to another without the requirement of a point source or focus position, we utilize Figure 7.26 to derive a vergence propagation relationship. The vergence V_1 in plane 1 corresponds to a focus at a distance L:

$$V_1 = \frac{n}{L} \tag{7.11 a}$$

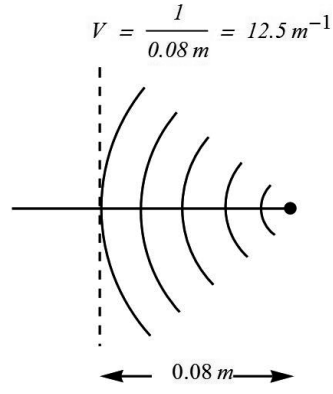


Figure 7.25 Converging example.

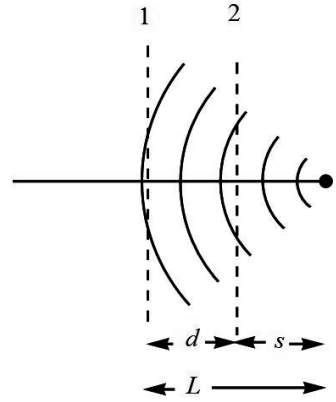


Figure 7.26 Vergence propagation.

Similarly, the vergence in plane 2 is

$$V_2 = \frac{n}{s} = \frac{n}{L-d} \tag{7.11b}$$

Substituting (7.11a) into (7.11b),

$$V_2 = \frac{n}{\frac{n}{V_1} - d} = \frac{V_1}{1 - V_1 \frac{d}{n}} \tag{7.11c}$$

Equation (7.11c) is known as the *vergence propagation* equation. The vergence propagation equation is useful for determining the vergence as a function of position. Consider the example of an expanding wavefront propagating through water (n = 1.33). The known wavefront has a vergence of –0.5 diopter. To find the vergence of a light wave 10m away from the known wavefront, use (7.11c) to get a vergence of –0.0105 diopter.

When a wavefront impinges on a lens or any surface with a different refractive index than the medium the wave is traveling in, the vergence changes. In Figure 7.27

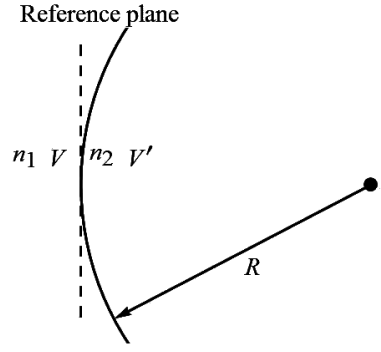


Figure 7.27 Vergence and surface power.

if V is the vergence of the wave to the left of the reference plane and V' is the vergence to the right of the reference plane, then

$$V' = V + \beta \tag{7.12}$$

where β is the power of the surface

$$\beta = \frac{n_2 - n_1}{R} \tag{7.13}$$

and, consistent with the convention we use for center of curvature, R is positive, indicating that if $n_2 > n_1$ the optical surface shown will try to convert a diverging wavefront to a converging wavefront and will convert a converging wavefront to a more rapidly converging wavefront.

The power of a lens in air is the reciprocal of the focal length[8]:

$$\begin{aligned} \beta_{lens} &= \frac{1}{f} \\ &= (n_l - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \end{aligned} \tag{7.14}$$

where (7.6b) was used to obtain the second equality. For a biconvex lens in air R_1 is positive and R_2 is negative. Assuming $|R_1| = |R_2| = R$, (7.14) becomes

$$\beta_{lens} = (n_l - 1)\frac{2}{R} \tag{7.15}$$

Example 7.3:
The image location and magnification of the system illustrated in Figure 7.28 is readily evaluated using the ray-tracing technique, Gauss's thin lens formula, and vergence concepts. An object is placed 10 cm to the left of a convex lens with a focal length of 5 cm. Using the ray-tracing technique, the image location is found to be 10 cm to the right of the lens. Using Gauss's law of (7.7e), we find s_i = 10 cm:

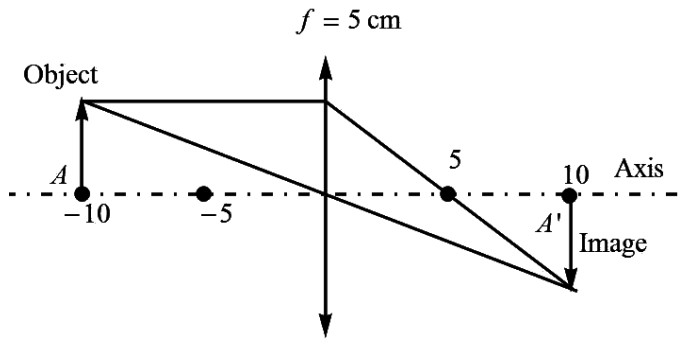


Figure 7.28 Ray tracing, Gauss's law, and vergence methods are used to find image location.

$$\frac{1}{-10}+\frac{1}{5}=\frac{1}{s_i} \quad \Rightarrow \quad s_i = 10\,\text{cm} \tag{7.16}$$

The vergence V_1 at the lens from the point located at A is 1/-0.1 = –10 diopters. The power of the lens β = 1/f = 1/0.05 = 20 diopters. The vergence V_1' just to the right of the lens is the sum of the vergence just to the left of the lens plus the power of the lens: V_1' = – 10 + 20 = 10 diopters. Referring to Figure 7.28, a vergence of 10 diopters corresponds to point A being imaged at point A' in Figure 4.28.

7.7 Multiple-Lens Systems

As illustrated in Figure 7.29, suppose two thin convex lenses with focal lengths of 10 and 15 cm, respectively, are separated by a distance of 35 cm. For an object located 30 cm away from the first lens, determine the location of the image using ray tracing, Gauss's equation, and vergence concepts.

The first lens is placed at the origin of the coordinate system. Then the second lens is at 35 cm, the focal points of the first lens are at coordinates – 10 and 10 cm, respectively; and the focal points of the second lens are at 20 and 50 cm, respectively. Using rays 1, 2, and 3 the image of P in the first lens is P_1. The image of P_1 is P_{12} and is found using rays 4, 5, and 6. The final system image is real, right side up, and is located at coordinate 95.

Now solve the problem using Gauss's thin lens equation:

$$\frac{1}{-30}+\frac{1}{10}=\frac{1}{s_{i1}} \quad \Rightarrow \quad s_{i1} = 15\,\text{cm} \tag{7.17a}$$

which is the location of P_1. Now change the coordinate system so lens 2 is at the origin and use Gauss's equation again:

$$\frac{1}{-20}+\frac{1}{15}=\frac{1}{s_{i2}} \quad \Rightarrow \quad s_{i2} = 60\,\text{cm to the right of lens 2} \tag{7.17b}$$

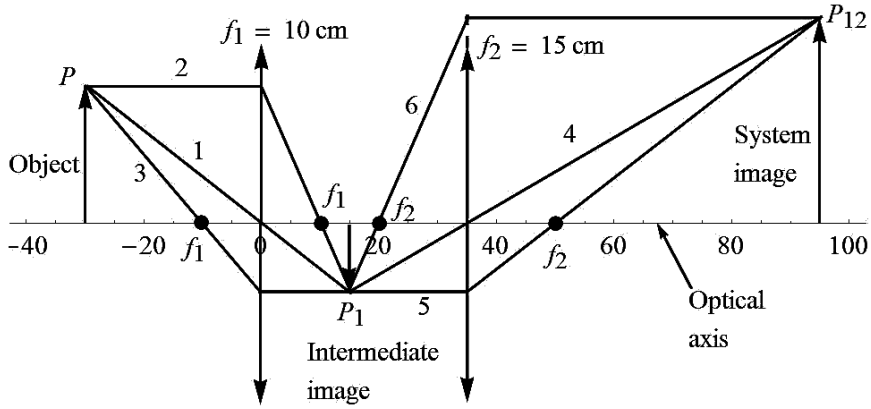


Figure 7.29 Multiple-lens system.

In the coordinate system of Figure 7.29, the image coordinate is 95.

Now solve the problem using the concept of vergence. The vergence V_1 just before the principal plane of the first lens is 1/ –0.30 = –3.333 diopters. The vergence V_1' just to the right of the principal plane of the first lens is –3.333 + 1/0.10 = 6.6667 diopters, which is positive, indicating that after the first lens the rays are converging. At the focal point the curvature will be infinite and the vergence will be zero. Solve (7.11c) for d:

$$\frac{6.6667}{1-6.6667d} = \infty \quad \Rightarrow \quad d = 0.15 \text{ m} \tag{7.18a}$$

The vergence V_2 just to the left of the principal plane for the second lens is 1/ –0.20 = –5 diopters. The vergence V_2' just to the right of principal plane for the second lens is –5 + 1/0.15 = 1.6667 diopters. Solve (7.11c) to determine how far away from lens 2 the vergence is infinite:

$$\frac{1.6667}{1-1.6667d} = \infty \quad \Rightarrow \quad d = 0.6\,\text{m} = 60\,\text{cm} \tag{7.18b}$$

Using the coordinate system of Figure 7.29, the image is formed 95 cm away from the first lens.

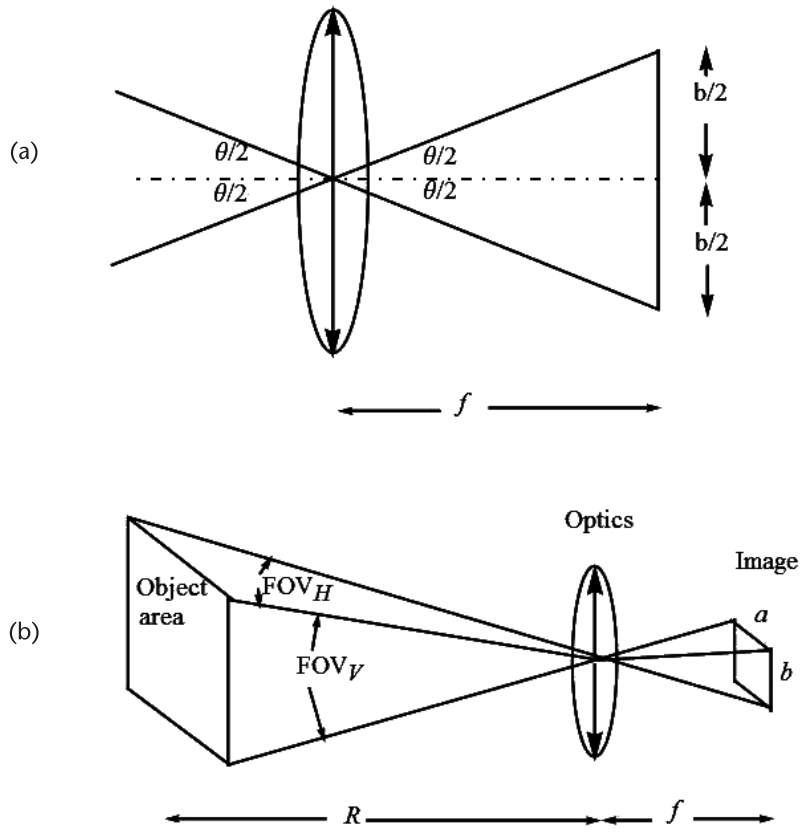


Figure 7.30 Field of view.

Gauss's equation and vergence concepts allow for a more accurate calculation of image location than the ray-tracing technique because of inherent inaccuracies in a graphical solution. However the ray-tracing technique provides guidance on the proper use of Gauss's equation and vergence concepts and is useful for understanding how an optical system works.

Alternate techniques for understanding optical systems include matrix methods[18,21,25] and ray-tracing calculations.[8,13,20] The Code V (http://www.opticalres.com) and Zeemax (http://www.zemax.com) optical design programs also do ray-tracing calculations. The treatment of these techniques is outside the scope of this text.

7.8 Field of View

The FOV of a FLIR or EO system is one of the most important design parameters since it describes the angular space in which the system accepts radiation. The system FOV rather than the magnification is specified because magnification depends strongly on object range, whereas for ranges long compared to a focal length, FOV is independent of target range. As illustrated in Figure 7.30, sensor FOV is determined by the size of the detector array and the system focal length.

From Figure 7.30(a):

$$\tan\frac{\theta}{2} = \frac{\frac{b}{2}}{f} \quad \Rightarrow \quad \theta = 2\tan^{-1}\frac{b}{2f} \quad \text{for distant objects} \tag{7.19a}$$

Applying this result to Figure 7.30(b),

$$FOV_H = 2\tan^{-1}\frac{a}{2f} \text{ and } FOV_v = 2\tan^{-1}\frac{b}{2f} \text{ for } R \gg f \tag{7.19b}$$

A limitation of (7.19) is that it assumes R is much greater than f so the image is a distance f from the principal plane. If the object of interest is at close range then (7.19a) and (7.19b) need to be modified. Let s_i denote the distance of the image from the principal plane. Then using Gauss's equation we can find s_i since R and f are known:

$$\frac{1}{R} + \frac{1}{s_i} = \frac{1}{f} \tag{7.20a}$$

In Figure 7.30, f gets replaced by s_i and the FOVs in the horizontal and vertical directions are, respectively,

$$FOV_H = 2\tan^{-1}\frac{a}{2s_i} \text{ and } FOV_v = 2\tan^{-1}\frac{b}{2s_i} \text{ for arbitrary } R \tag{7.20b}$$

A *field stop* is an aperture at an image plane (or an intermediate image plane) of an optical system that determines the size and shape of the image.[22] In Figure 7.30,

a and *b* determine the field stop. In an imaging system that uses film, the exposed portion of the film determines the field stop. FLIR and EO systems use detectors to form the image. The detectors take the form of a two-dimensional array of individual detectors called a staring array or a single detector or rows of detectors that are scanned across the image space. In these cases, the size of the image plane is defined by the light-sensitive area of the array and the possible light-sensitive positions of the scanned system.

The optics of a system usually comprises a number of lenses with a single *effective focal length* (sometimes shortened to just *focal length*). This is the equivalent focal length of the system and is illustrated in Figure 7.31 for a converging optical system. The incident collimated rays are projected forward and the rays leaving the optical system are projected backward. The rays meet at a principal plane. The distance from this plane to focal point *f* is the effective focal length of the system. Here we have assumed the refractive index n_m in the object space is the same as the refractive index in the image space. If this is not the case, then the optical system has an effective focal length in image space and a different effective focal length in object space and these are found using a technique similar to that used in Figure 7.31. The effective focal length of a diverging optical system is defined by a similar procedure.

Figure 7.32 illustrates how to calculate the effective focal length of an optical system. Collimated rays parallel to the optical axis and incident on the optical system come to a focus at coordinate labeled P' in the figure. Using Gauss's thin lens equation, we find the rays from P' are brought to a focus at P'', which is 7.5 cm to the right of the 5-cm focal length lens. Suppose each incident collimated ray is a distance *h* away from the optical axis. Then the collimated rays intersect the 5-cm focal lens at a height $h/2$. Back tracing the ray from P'', we see that in going 7.5 cm the ray has moved a distance $h/2$ away from the optical axis, which implies that in a horizontal distance of 15 cm the backward-traced ray will intersect the forward-traced ray. This implies that for the purpose of calculating image size and FOV the two-lens system illustrated in Figure 7.32 can be replaced by a single thin lens with a focal length of 15 cm placed 15 cm to the left of point P''.

The effective focal length concept is useful for calculating the FOV for a multiple-lens system. Suppose a detector array 1 cm in size is placed at point P''. Then the FOV of the system is readily calculated using (7.19b):

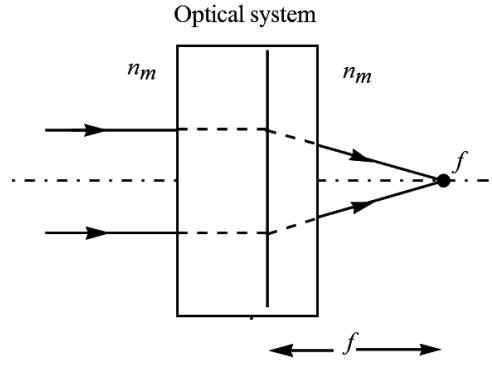


Figure 7.31 Effective focal length for a converging optical system.

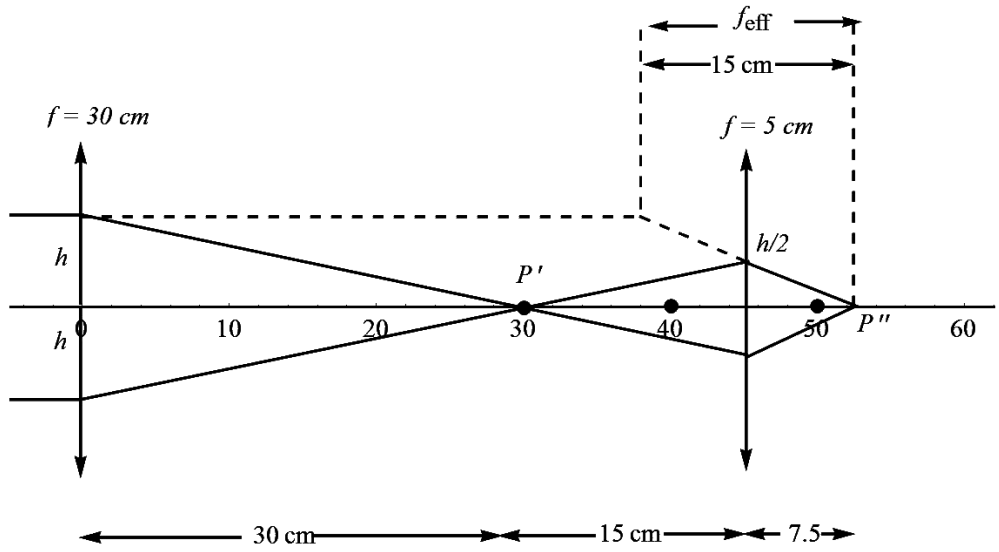


Figure 7.32 Effective focal length example.

$$FOV_v = 2\,tan^{-1}\frac{b}{2f} = 2\,tan^{-1}\frac{1}{2\times 15} = 0.0666\,\text{rad} = 66.6\,\text{mrad} \tag{7.21a}$$

We now illustrate a technique for calculating the FOV of a multiple-lens system without first calculating the effective focal length. Since P'' and P' are conjugate points, a 1-cm-high detector array placed at P'' will be imaged as a 2-cm-high detector array at P'. Then the system FOV is calculated using (7.19b) for the 30-cm lens:

$$FOV_v = 2\ tan^{-1}\frac{b}{2f} = 2\ tan^{-1}\frac{2}{2\times 30} = 0.0666\,\text{rad} = 66.6\,\text{mrad} \tag{7.21b}$$

which is in agreement with the effective focal length approach.

Now we discuss the design nature of the FOV. A wide FOV allows the sensor to view a larger area, but the detector elements are spread over this larger area, yielding lower resolution (i.e., image fidelity). A narrow FOV gives better resolution but it is difficult to find objects of interest over a large area. Military applications require the sensor to have at least two FOVs. The first is a medium FOV for finding or detecting objects of interest, and the second is a narrow FOV that can be used to interrogate potential targets. For tactical systems, large FOVs are about 10 deg and up, medium FOVs are around 5 deg, and narrow FOVs are less than 3 deg.

The trade-offs for system FOV include sensor resolution, sensitivity, and area coverage. The optics and detectors are designed to provide sufficient resolution and sensitivity for a given task. Overall sensor performance is evaluated to accomplish discrimination of a particular target and finally the largest possible FOV is selected that still allows the required target discrimination. In Chapter 10 we describe the procedure for accomplishing both discrimination performance and FOV selection.

7.9 Resolution

The concept of diffraction was presented in Chapter 4 so we know that a lens cannot provide a perfect image of a point source. Even for perfect lenses with no aberrations (imaging imperfections caused by refraction or reflection), the images of point objects are not points but are optical psfs. With or without aberrations, there is an ultimate resolving limitation for a lens or optical system. The resolution of a lens or optical system can be specified by either Rayleigh or Sparrow's criterion.[9]

If we consider the geometrical optics representation, light is characterized as traveling in straight lines. The ray representation of light from an ideal point source at infinity and imaged by a lens is shown in Figure 7.33.

From the approach of a perfect imaging system, the point source would be focused to an ideal point and hence would produce infinite resolution. This is not the case for real-world systems. The actual size of the imaged point source is determined by three contributions: (1) diffraction, (2) aberrations, and (3) manufacturing defects and assembly tolerances. The actual image of a point source by a real system is often referred to as a *blur circle*. For many optical systems, diffraction produces the primary constraint on resolution because aberrations and manufacturing defects can be minimized by good optical design and tight manufacturing tolerances and alignment procedures. Figure 7.34 illustrates the relative spatial intensity of a diffraction-limited imaged point in dimensionless radial coordinates.

The intensity profile of a point-source image for a diffraction-limited optical system is given by[7,10]

$$I(r) = I_0\left(\frac{2J_1\left(ka\frac{r}{f}\right)}{ka\frac{r}{f}}\right)^2 = I_0\left(\frac{2J_1\left(\pi D\frac{r}{f\lambda}\right)}{\pi D\frac{r}{f\lambda}}\right)^2 = I_0\,somb^2\left(D\frac{r}{f\lambda}\right) \tag{7.22}$$

where r is the radial distance from the peak, a is the lens radius and is equal to $D/2$, f is the lens focal length, $k = 2\pi/\lambda$, and I_0 is the value of $I(r)$ when r is zero. With I_0 set to one, this function is the psf of the optics.

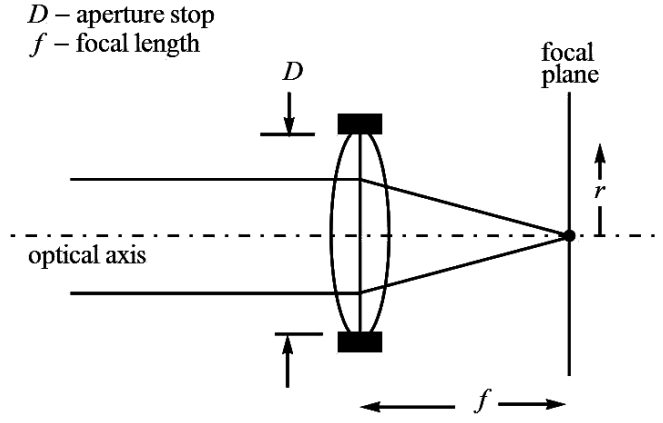


Figure 7.33 Perfect point imaging system.

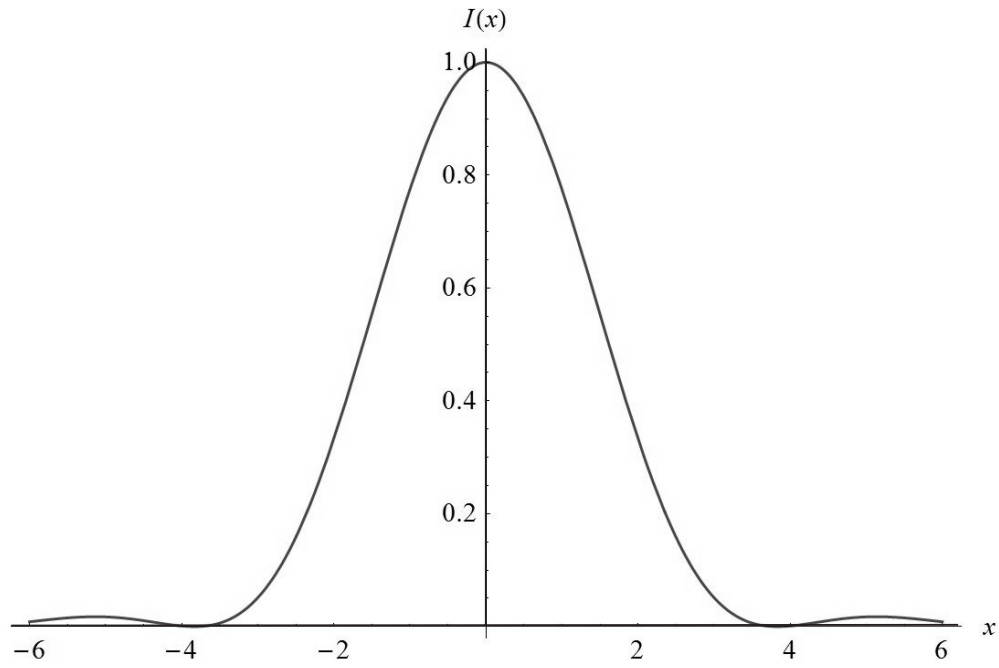


Figure 7.34 Cross section of diffraction-limited point intensity.

Given that a point source is not imaged as a perfect point but is spread out in the image plane, we want to be able to define just how well an optical system can resolve objects that are placed close together. In other words, what is the spatial resolution of the sensor system? Consider two separate, adjacent, equal-intensity point sources that are imaged very close together, as shown in Figure 7.35.

The Rayleigh resolution limit defines two point sources as just resolved when the peak of one point-source image coincides with the first zero of the second

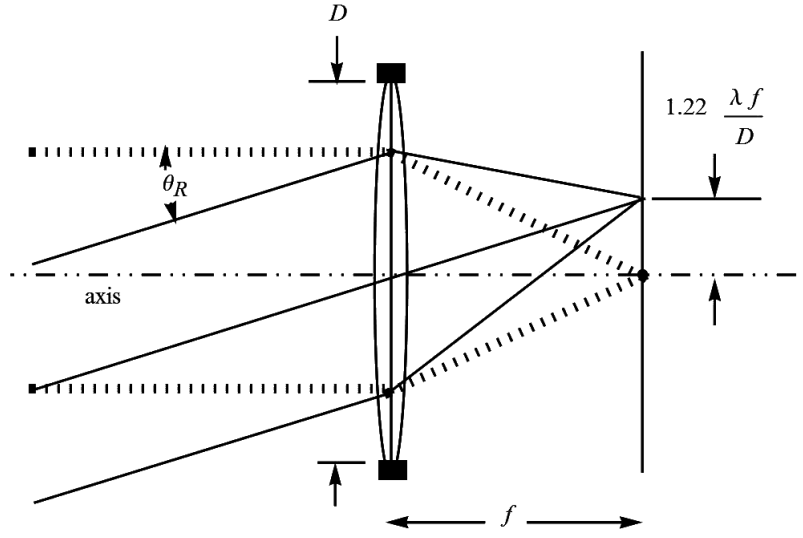


Figure 7.35 System imaging two point sources.

point-source image. As seen in Figure 7.34, the first zero of the $J_1(x)$ is at 3.83. From (7.22), this implies that the radial coordinate of the first zero is $(3.83/\pi)$ $(\lambda f/D)$ = $1.22\ \lambda f/D$ as illustrated in Figure 7.35. Expressing this as an equation, we get

$$r_R = 1.22\frac{\lambda f}{D} \tag{7.23}$$

where r_R is a distance in the image plane. Another way to express (7.23) is in angular coordinates:

$$\theta_R = 1.22\frac{\lambda}{D} \tag{7.24}$$

where θ_R is the angular distance between perfect point sources at an infinite distance from the lens. Note that the Rayleigh criterion is not a mathematical law but a rule of thumb that is commonly used to define the diffraction-limited resolution of an optical sensor. Figure 7.36 illustrates the intensity profile in the image plane of two point sources separated by $1.22\ \lambda f/D$.

Another common technique for defining optical resolution is Sparrow's criterion, which states "two image points A' and B' in the image plane can be just resolved if the intensity is a constant as we move from the image point A' to the point B' in the image plane." The Sparrow angular resolution is given by

$$\theta_s = \frac{\lambda}{D} \tag{7.25}$$

The Sparrow angular separation is less than that of the Rayleigh criterion. Both of these criteria, Rayleigh and Sparrow, were derived assuming incoherent sources.

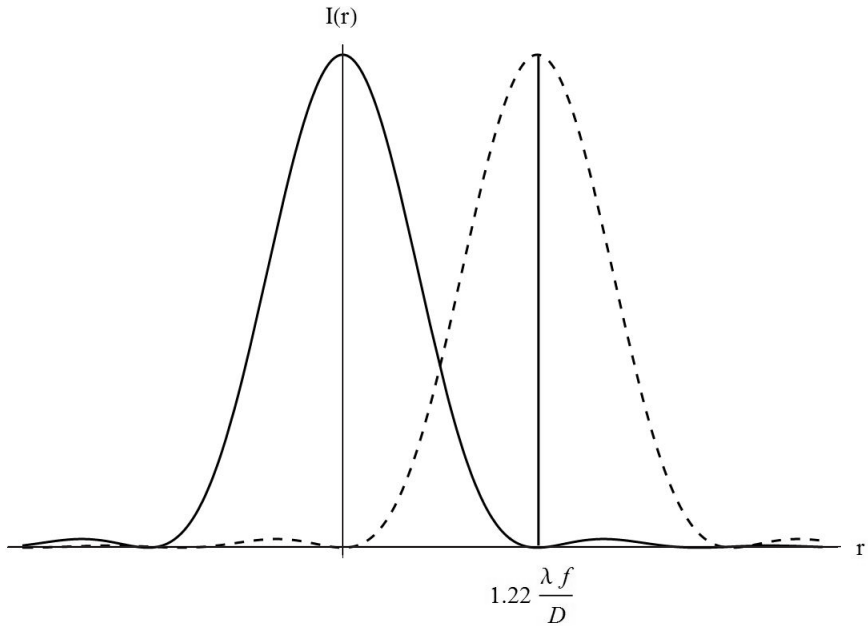


Figure 7.36 Intensity of two point sources just resolved by Rayleigh limit.

Another common term for defining resolution is the *blur circle* or the *Airy disk*. This definition uses the same Bessel-based *sombrero* intensity equation defined above but defines the central blur as containing 84% of the energy. The angular size of the blur is given by

$$\theta_{Airy\,disk} = 2.44\frac{\lambda}{D} \tag{7.26}$$

The diameter of the blur in the image plane is given by

$$d_{Airy\,disk} = 2.44\frac{\lambda f}{D} \tag{7.27}$$

Because the *f*/ # (*f*-number) is approximately *f*/*D*, the spot size can be thought of as a function of *f*/ # rather than as a function of optics diameter and focal length. The overall goal in high-resolution sensor design is for the system to be able to resolve two closely separated sources, which equates to minimizing $\theta_{Airy\,disk}$. The final design needs to consider detector characteristics, FOV, and sensitivity requirements.

7.10 Aperture Stop, Pupils, and Rays

A knowledge of stop, pupil, and ray nomenclature is important for understanding optical systems. The aperture stop is important because it determines the irradiance in the image plane and is useful for predicting the SNR. We have already seen that the field stop determines the FOV of an optical system. Our treatment of diffraction up to this point has been for a distant object and a thin lens. To extend the theory to near objects and thick lenses, it is helpful to understand entrance and exit pupils. Pupils are also important in combining two optical subsystems to make an optical system. Good optical design requires the exit pupil of the first optical subsystem to coincide in position and size with the entrance pupil of the second optical subsystem. An understanding of ray nomenclature and the use of different rays facilitates an understanding of optical systems.

Several optical concepts are illustrated in Figure 7.37. The *aperture stop* is the surface in the optical system that limits the angle over which the optical system will accept rays from an on-axis object point. An *axial ray* is a ray that begins at an on-axis point in the object plane. An axial ray that is always close to the optical axis is called a *paraxial ray*. The paraxial ray is approximately normal to all refracting and reflecting surfaces and this implies that the angle the paraxial ray makes with the normal is small. If the paraxial ray is sufficiently close to the optical axis, the sine and tangent of the reflection or refraction angle can be approximated by the angle, and the cosine of the angle can be approximated by one. This simplifies optical calculations and is the reason paraxial rays are important. A useful property of the paraxial ray is that if it originated on the object, an intermediate image appears wherever the paraxial ray crosses the optical axis. A *marginal ray* is an axial ray whose angle with respect to the optical axis is large enough so that it barely makes it through the aperture stop. One use of the marginal ray is to determine the aperture

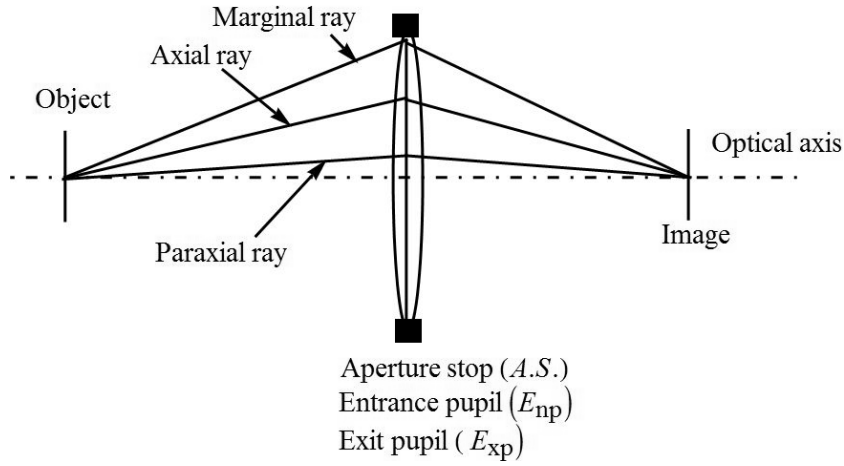


Figure 7.37 Aperture stop, entrance pupil, and exit pupil for a thin lens. Also shown are paraxial, axial, and marginal rays.

stop of the optical system. The *entrance pupil* is the image of the aperture stop in the optical elements between the aperture stop and the object. The entrance pupil is what is seen by an axial point in the object plane when viewing the aperture stop. The *exit pupil* is the image of the aperture stop in the optical elements between the aperture stop and the image.

From Gauss's thin lens equation, we deduce that for the thin lens the aperture stop, the entrance pupil, and the exit pupil coincide and this is consistent with Figure 7.38. In more complicated optical systems, the marginal rays can be used to determine the extent of the entrance and exit pupils.

The *chief ray* (sometimes called the *principal ray*) is defined as a ray that starts from an extreme off-axis point on the object and goes through the center of the aperture stop. The chief ray either goes through the center of the entrance and exit pupils as illustrated in Figure 7.38 or extensions of the chief ray go through the center of the entrance and exit pupils. The reason for this is that the entrance and

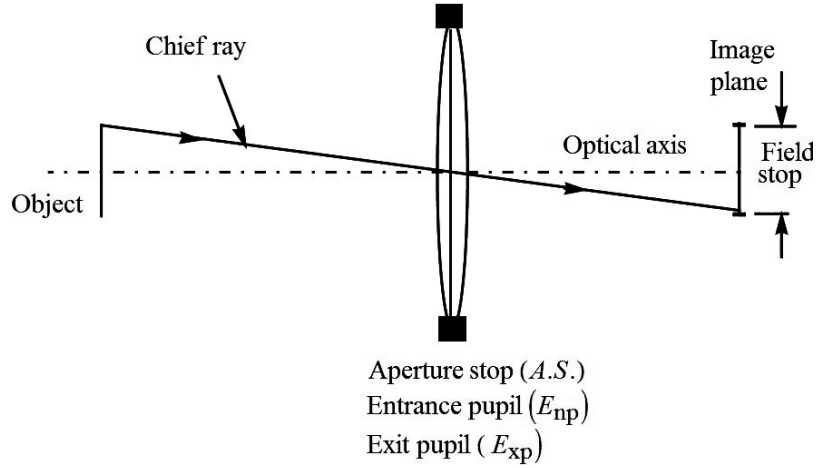


Figure 7.38 Illustration of the chief ray and field stop for a thin lens.

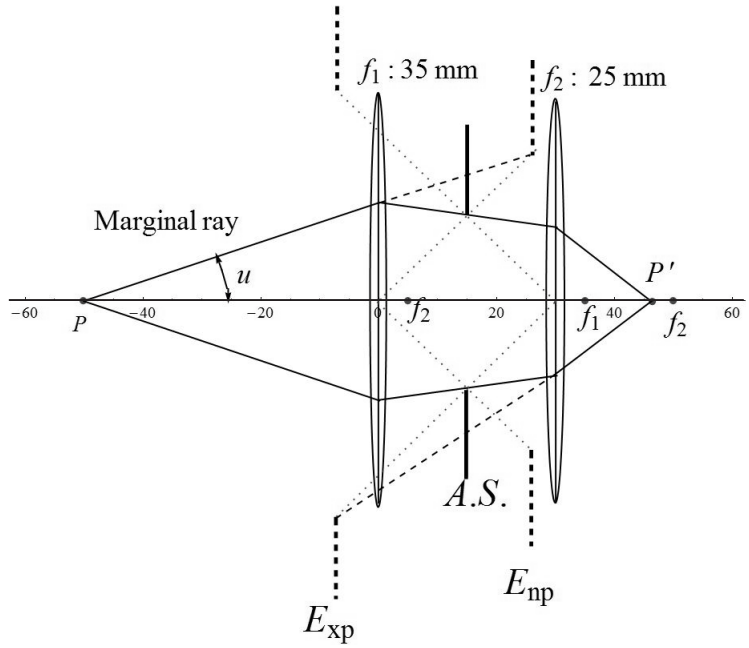


Figure 7.39 Ray trace for determining the aperture stops and entrance and exit pupils.

exit pupils are images of the aperture stop. The *field stop* is an aperture in an image plane that determines the size and shape of the image.[22]

Figure 7.39 illustrates one technique for finding the aperture stop for an axial point P. The angle u is increased and the ray from P is traced to determine what prevents it from getting through the optical system. Here the rays were traced using the ray-tracing rules of Section 7.3 with some help from Gauss's thin lens equation. In this case the two lenses are sufficiently large and it is the baffle (labeled *A.S.* in the figure) that determines the marginal rays. The entrance pupil is the image of the aperture stop in the 35-mm focal length lens and this is indicated by the dotted lines emanating from the center of this lens. The exit pupil is the image of the aperture stop in the 25-mm focal length lens and this is indicated by the dotted lines emanating from the center of this lens. Note that extensions of the marginal rays from P and P' (denoted in the figure by dashed lines) intersect the boundaries of entrance and exit pupils.

Having determined the aperture stop, entrance and exit pupil in Figure 7.39, in Figure 7.40 the chief ray is traced. Note that the chief ray goes through the center of the aperture stop. When the chief ray leaves point Q it is heading toward the center of the entrance pupil, which makes sense since the entrance pupil is the image of the aperture stop in the 35-mm focal length lens. Similarly, the ray arriving at point Q' appears to come from the center of the exit pupil. In Figures 7.39 and 7.40, the entrance and exit pupils are represented by dashed rather than solid lines to indicate that the entrance and exit pupils are virtual.

In Figure 7.39 we showed how to find the location and size of the entrance and exit pupil using ray-tracing rules and Gauss's thin lens equation. Another way to find the entrance and exit pupil locations is to start a ray at the center of the aperture stop (by definition this is a chief ray) and trace it backward and forward

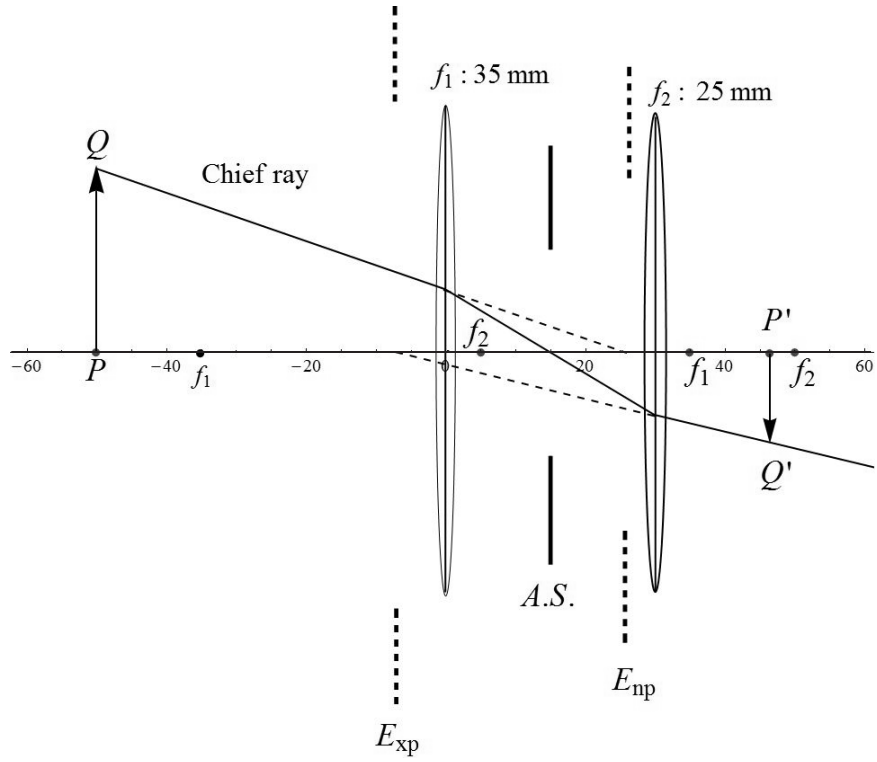


Figure 7.40 Ray trace showing relationship of chief ray to aperture stop and entrance and exit pupils.

through the optical system.[8] Where the ray crosses the optical axis (or appears to cross the optical axis as in Figure 7.40) determines the location of the entrance and exit pupils, which are real or virtual depending on whether the ray actually crossed the optical axis or an extension of the ray crossed the optical axis. Having determined the location of the pupils, the extremities of the marginal rays at the location of the entrance and exit pupils determine the size of the entrance and exit pupils.

In Figure 7.39 the baffle between the two lenses was the aperture stop. However, if the 35-mm or 25-mm lens in that diagram were made smaller, then one of them could become the aperture stop. The easiest way to determine the aperture stop is to do a meridional ray trace from a point *P* on the optical axis to determine what element defines the marginal ray. Once the aperture stop is determined, the entrance pupil is the image of the aperture stop using optical elements between the aperture stop and the object. Similarly, the exit pupil is the image of the aperture stop using optical elements between the aperture stop and the image.

7.11 The f-Number and Numerical Aperture

The *f*-number and *numerical aperture* are useful concepts for estimating the power per unit area in the image given the luminance of the imaged object. The *f*-number, often represented by the symbol *f*/#, is used to describe the SNR of EO systems, to describe the size of the diffraction blur spot in object or image space, and to

characterize aberrations (see rules of thumb and *RCA Handbook*). The *f*-number is sometimes referred[11,26] to as the *focal ratio*, *f-ratio*, *f-stop*, and *relative aperture*. There are different definitions for *f*-number, each appropriate for a particular use.

Object space is all of the space to the left of the first element of the optical system; *image space* is the space to the right of the last optical element.[20] Here it is assumed that the object is to the left of the optical system and light travels from left to right. In Figure 7.41, n and n' denote the refractive index in object and image space, respectively; P is an axial point in the object plane and P' is the corresponding point in the image plane. The angle the marginal ray makes with the optical axis at P and P' are denoted by u and u'. Then the *numerical aperture* is defined[9] by the relation

$$NA \equiv n' \sin u' \tag{7.28}$$

The numerical aperture is measured with respect to a particular object or image point and will vary as that point is moved. Thus, as point P is moved, the numerical aperture will change because the angles u and u' change. The numerical aperture also depends on the size of the aperture stop and gets larger as the aperture stop gets larger. For systems corrected for coma and spherical aberration and with the object being at infinity the *f*-number and numerical aperture are related[9,26]

$$f/\# \equiv \frac{1}{2\,NA} = \frac{1}{2n' \sin u'} \tag{7.29a}$$

An alternate definition for the *f*-number of an optical system is[5,8,20,22]

$$f/\# \equiv \frac{f_{eff}}{D_{enp}} \tag{7.30a}$$

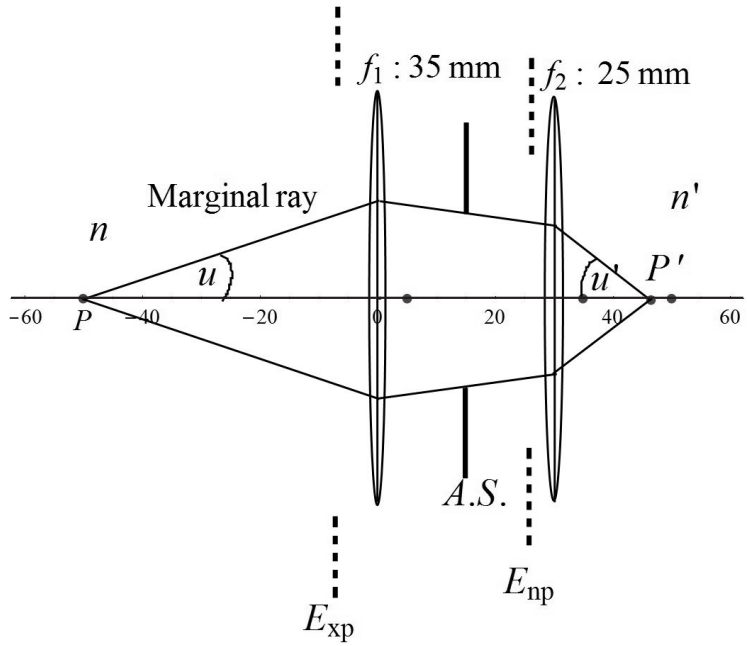


Figure 7.41 Numerical aperture and *f*-number.

where f_{eff} and D_{enp} are the effective focal length and diameter of the entrance pupil, respectively, measured in the same units of length so that the *f*-number is dimensionless. Note from Figure 7.41 that in general the diameter of the entrance pupil is *not* equal to the diameter of the first optical element. The technique for finding the diameter of the entrance pupil is given in Section 7.10 and the technique for finding the effective focal length is given in Figure 7.32. Note from Figure 7.32 that in general f_{eff} is not the focal length of any particular lens. As illustrated in Figure 7.42, for a thin lens the diameter of the entrance pupil is the clear diameter of the lens and the effective focal length is measured from the thin lens principal plane.

Referring to Figure 7.42, (7.29a) becomes

$$f/\# = \frac{1}{2\sin u'} \tag{7.29b}$$

while (7.30a) becomes

$$f/\# = \frac{1}{\frac{D_{enp}}{f_{eff}}} = \frac{1}{2\frac{\frac{D_{enp}}{2}}{f_{eff}}} = \frac{1}{2\tan u'} \tag{7.30b}$$

Since the radiometry at point P' depends sensitively [see (7.47)] on the angle u', solve (7.29b) and (7.30b) or the angle u' in terms of the of *f*/# using the following two definitions:

$$u' = \sin^{-1}\left(\frac{1}{2f/\#}\right) \tag{7.29c}$$

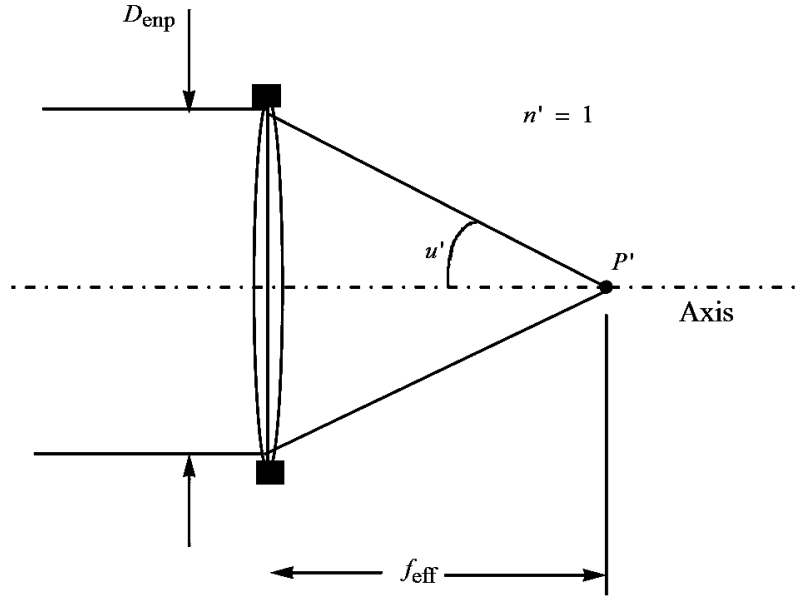


Figure 7.42 Relationship between *NA* and *f*/# for a thin convex lens in air.

$$u' = tan^{-1}\left(\frac{1}{2f/\#}\right) \tag{7.30c}$$

Figure 7.43 shows that for $f/\# < 1$ the two definitions give significantly different u' angles but for most purposes the difference in the u' angle will be negligible for $f/\# > 1.5$. The power series expansion for $sin^{-1}\, x$ and $tan^{-1}\, x$

$$sin^{-1}x = x + \frac{x^3}{6} + \frac{3x^5}{40} + \cdots \tag{7.31}$$

$$tan^{-1}x = x - \frac{x^3}{3} + \frac{x^5}{5} + \cdots \tag{7.32}$$

demonstrates that for small x (corresponding to large $f/\#$) the two definitions of (7.29a) and (7.30a) are equivalent. Equation (7.30a) is the most often used definition of f-number.

We will now show that the power on a detector from an extended source is inversely proportional to the square of the optical system f-number. With τ_{atm} denoting the transmittance of the atmosphere, the flux incident on the optical system is

$$\Phi_{optics} = L_s A_s \frac{A_{optics}}{s_o^2} \tau_{atm} \tag{7.33}$$

The flux in the image is the flux incident on the optical system reduced by the transmission of the optics

$$\Phi_{image} = \Phi_{optics}\tau_{optics} = L_s A_s \frac{A_{optics}}{S_o^2} \tau_{atm}\tau_{optics} \tag{7.34}$$

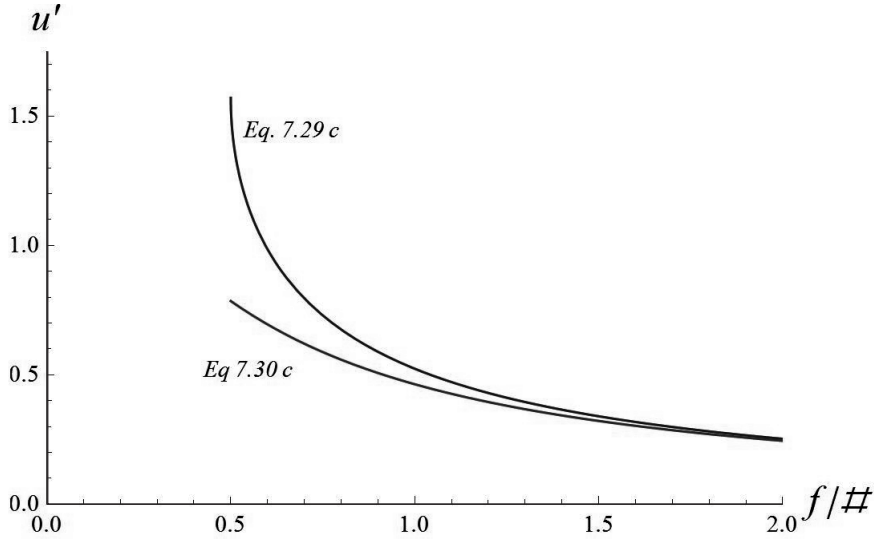


Figure 7.43 Demonstrating inconsistencies in accepted definitions of $f/\#$.

Because the source is characterized by a uniform radiance, the image also has a uniform radiance, which implies the flux on the detector is given by

$$\Phi_{det} = \Phi_{image} \frac{A_d}{A_i} = L_s A_s \frac{A_{optics}}{S_o^2} \frac{A_d}{A_i} \tau_{atm} \tau_{optics} \tag{7.35}$$

The ratio of the source area to the image area is

$$\frac{A_s}{A_i} = \frac{S_o^2}{S_i^2} \quad \Rightarrow \quad \frac{1}{A_i} = \frac{S_o^2}{S_i^2} \frac{1}{A_s} \tag{7.36}$$

Using this result (7.35) becomes

$$\Phi_{det} = L_s A_s \frac{A_{optics}}{S_o^2} \frac{A_d}{A_s} \frac{S_o^2}{S_i^2} \tau_{atm} \tau_{optics} = L_s A_{optics} A_d \frac{1}{S_i^2} \tau_{atm} \tau_{optics} \tag{7.37}$$

This shows that for an extended source the flux on the detector is independent of the source size. We want to express (7.37) in terms of the focal length f and the magnification M. Using Gauss's equation and the magnification relationship for a thin lens

$$\frac{1}{S_o} + \frac{1}{S_i} = \frac{1}{f}, \quad M = \frac{S_i}{S_o} \tag{7.38}$$

Gauss's equation can be written

$$\frac{M}{S_i} + \frac{1}{S_i} = \frac{1}{f} \quad \Rightarrow \quad \frac{1}{S_i} = \frac{1}{f(1+M)} \tag{7.39}$$

Using this relationship, (7.37) becomes

$$\Phi_{det} = \frac{1}{f^2 (1+M)^2} L_s A_{optics} A_d \tau_{atm} \tau_{optics} \tag{7.40}$$

Expressing A_{optics} in terms of D_{optics}, the diameter of the thin lens depicted in Figure 7.44, we find

$$\begin{aligned} \Phi_{det} &= \frac{\pi}{4} D_{optics}^2 \frac{1}{f^2 (1+M)^2} L_s A_d \tau_{atm} \ \tau_{optics} \\ &= \frac{\pi}{4} \frac{1}{f/\#^2 (1+M)^2} L_s A_d \tau_{atm} \ \tau_{optics} \end{aligned} \tag{7.41}$$

where we have used the definition (7.30a) to get the second equality. Equation (7.41) demonstrates that the flux on a detector is proportional to the source radiance and the detector area and inversely proportional to the square of the f – number. For a distant object, $s_i \approx f$ and we see from (7.38) that in this case the factor $M \approx 0$. If the object is at a distance twice that of the focal length from the focal plane, then

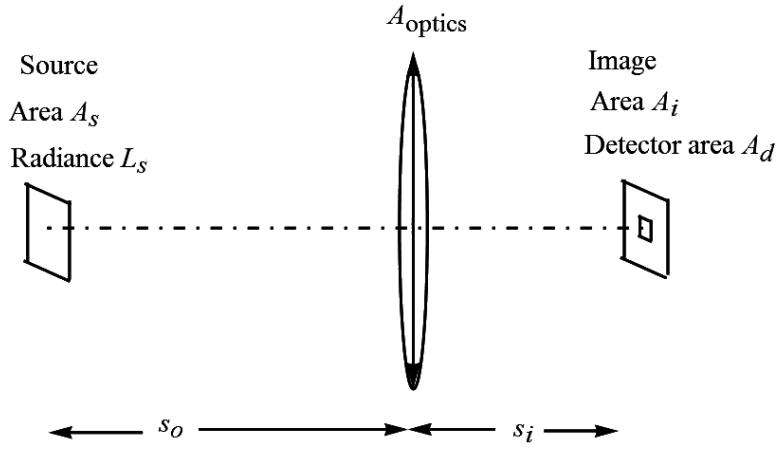


Figure 7.44 Radiometry of an extended source.

s_i equals s_o and $M = 1$. Using (7.41) we observe that, ignoring transmission losses, the flux on the detector is reduced by a factor of 4 compared to the case where the object is far away. This result, which some readers may find counterintuitive, can be understood in the following way. When the object is close, the wave incident on the optic has a high curvature and the lens tries to bend the rays back to the optical axis but because the incident wave has a high curvature the image is formed at a distance of $2f$ from the lens rather than at a distance of f. Because the image is further from the lens, it is larger than the image when the object is distant and for that reason is dimmer.

An interesting and useful result is that, aside from transmission losses, the radiance of the object and the radiance in the image are the same. This is easily seen by dividing both sides of (7.37) by $A_d A_{optics} / S_i^2$:

$$\frac{\Phi_{det}}{A_d \dfrac{A_{optics}}{S_i^2}} = L_s \tau_{atm} \tau_{optics} \quad \Rightarrow \quad L_i = L_s \tau_{atm} \tau_{optics} \tag{7.42a}$$

where L_i is the radiance of the image. Although we derived (7.42a) for a simple lens, the result applies to any optical system as long as the refractive index in image and object space is the same. If the refractive index is not the same in image and object space, then (7.42a) needs to be modified:

$$\frac{L_i}{n_i^2} = \frac{L_s}{n_s^2} \tau_{atm} \tau_{optics} \tag{7.42b}$$

The working f-number, denoted by the symbol $f/\#_W$, is defined by the relationship[8]

$$f/\#_W \equiv (1 + M) f/\# \tag{7.43}$$

One value of the working f-number is that it enables (7.41) to be expressed more simply. The working f-number is always greater than or equal to the f-number. This

reflects the fact that close-up objects are imaged at a greater distance than a focal length from the principal plane. Some authors[8] place a minus sign before the magnification, which indicates whether or not the image is inverted. If this is the case, then the absolute value of M should be used in (7.43).

Equation (7.41) shows that to maximize the flux on a detector it is desirable to minimize f-number and this suggests the question: *How small can the f-number be*? From the definition of (7.30a), it would seem that we could make the f-number arbitrarily small by using a highly refractive medium with high curvature to make f arbitrarily small. Examining the definition (7.29b), we see that the minimum value of f-number corresponds to the value of u', which maximizes $\sin(u')$. Using the definition (7.29b), the conceptual minimum f-number is $\frac{1}{2}$. As indicated by (7.47), the definition (7.29b) is appropriate for calculating the flux on a detector when the angle u' is large. We conclude that he smallest allowed value for the $f/\#$ in (7.41) is ½ in agreement with Holst.[11] In practice, one rarely encounters an EO system with an f-number of less than 1 and this can be considered a practical lower limit to the f-number.

Equation (7.41) is useful for finding the flux on a detector from an extended source if the f-number and magnification are known. An alternate formulation expresses the flux on the detector in terms of the half angle θ_{max} ubtended by a circular extended source. The irradiance E_d on a detector from an extended circular source with uniform radiance L_s is[6]

$$E_d = \pi\, L_s \sin^2 \theta_{\max} \tag{7.44}$$

We now derive this result. Using the definition of radiance and Figure 7.45, the flux on a differential element of detector area from a differential element of the source is written

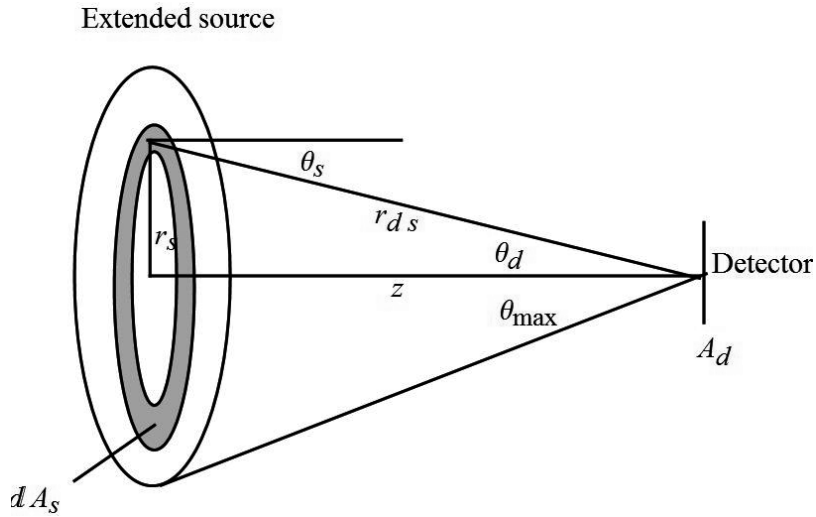


Figure 7.45 Flux on a detector from an extended source.

$$\Delta^2\Phi_d = L_s \Delta A_s \cos\theta_s \Delta\Omega_d = L_s \Delta A_s \cos\theta_s \frac{\Delta A_d \cos\theta_d}{r_{ds}^2} \tag{7.45}$$

The flux on the detector is found by integrating over the detector area and the source

$$\Phi_d = \int_{A_d}\int_{A_s} L_s \, dA_s \cos\theta_s \frac{\cos\theta_d}{r_{ds}^2} dA_d \tag{7.46}$$

To evaluate the integral realize that

$$\theta_d = \theta_s \equiv \theta; \; r_{ds} = \frac{z}{\cos\theta}$$

$$r_s = z\tan\theta \quad \Rightarrow \quad dr_s = \frac{z}{\cos^2\theta} d\theta$$

$$dA_s = 2\pi r_s \, dr_s = 2\pi z \tan\theta \frac{z}{\cos^2\theta} d\theta$$

Using these results the flux on the detector is readily calculated:

$$\Phi_d = \int_{A_d}\int_{\theta} L_s \, 2\pi z \tan\theta \frac{z}{\cos^2\theta} d\theta \cos^2\theta \left(\frac{z}{\cos\theta}\right)^{-2} dA_d$$

$$= 2\pi L_s A_d \int_{\theta=0}^{\theta_{\max}} \sin\theta\cos\theta \, d\theta = 2\pi L_s A_d \frac{1}{2}\sin\theta_{\max} \tag{7.47}$$

Divide the flux on the detector by A_d to get (7.44). In deriving the result of (7.44), we used an extended circular source with uniform radiance. Suppose now that the circular source in Figure 7.45 is a lens and to the left of the lens is an on-axis extended source with radiance L_s and the detector is positioned at the z coordinate so the extended source is in focus. To the detector, the lens will appear to be a uniform source with radiance L_s and so (7.44) applies to this case as well.

What happens to the irradiance if the object is away from the optical axis? Using Figure 7.46, we will show that the irradiance falls off as $\cos^4\theta$. Using the definition of radiance, the flux Φ_d on the detector is

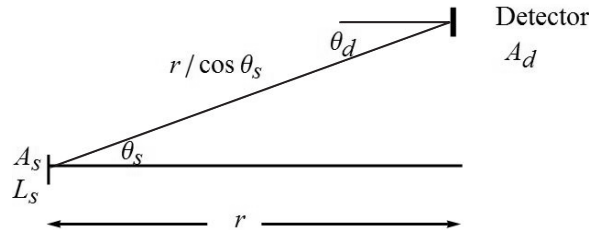


Figure 7.46 Irradiance of a detector by a point source.

$$\Phi_d = L_s\, A_s \cos\theta_s \frac{A_d \cos\theta_d}{\left(r/\cos\theta_s\right)^2} = \frac{L_s\, A_s\, A_d}{r^2} \cos^3\theta_s \cos\theta_d = \frac{L_s\, A_s\, A_d}{r^2} \cos^4\theta \qquad (7.48)$$

where we have made the identification that $\theta_s = \theta_d \equiv \theta$. The irradiance on the detector is obtained by dividing both sides of this equation by A_d:

$$E_d = \frac{L_s A_s}{r^2} \cos^4\theta = E_0 \cos^4\theta \qquad (7.49)$$

where E_0 is the irradiance the detector would see when θ is zero. Although we have derived this equation without regard to an optical system, (7.49) holds for an optical system if we identify θ with the angle between the optical axis and a line connecting the detector and the center of a thin lens.

Vignetting can further reduce the irradiance the detector sees. The aperture stop defines the marginal rays for an on-axis point. In some optical systems, for off-axis points, the aperture stop and some other optical element[20] determine the rays that make it through the system and effectively reduce the size of the aperture stop. This is called *vignetting*. Vignetting occurs[26] when there is a mismatch between elements of an optical system and this produces a partial obstruction of the aperture for some points in the image plane. For optical systems that have vignetting, (7.49) is valid for angles below some critical value θ_c (θ_c is the angle between the optical axis and a ray drawn from an image point to the center of the lens) and above that angle there is a more rapid drop-off in irradiance that sometimes results in a noticeable darkening at the periphery of the image. Sometimes, vignetting is intentionally introduced into an optical system to eliminate aberrated rays at the expense of flux throughput. Vignetting is much more common in wide FOV systems as opposed to narrow FOV systems such as telescopes. A schematic example of an optical system with a vignetted field is shown in Figure 7.47.

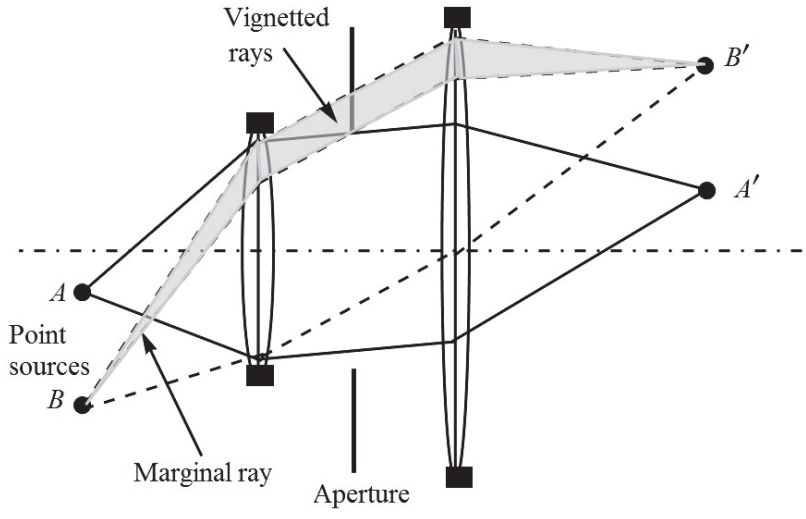


Figure 7.47 Vignetting in an optical system.

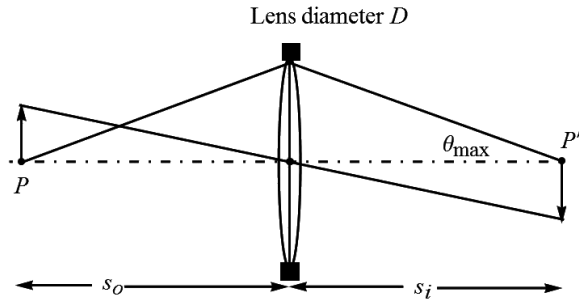


Figure 7.48 Determining $\theta_{\max}$.

The bundle of rays from point source A passes through the aperture and the two lenses and are imaged at A'. However, the rays from point source B are partially blocked by the aperture and the lenses reducing the irradiance at point B'.

To use (7.44) when an object is imaged by a lens, we need to be able to calculate $\theta_{\max}$. Referring to Figure 7.48, the image distance s_i is readily found from the object distance s_o using the lens focal length and Gauss's thin lens equation. Then the sine the angle the marginal ray makes with the optical axis is

$$sin\,\theta_{\max} = \frac{\frac{D}{2}}{\sqrt{s_i^2 + \left(\frac{D}{2}\right)^2}} \tag{7.50}$$

which enables the irradiance at the point P' to be found using (7.44). For a thick lens the equation is the same but s_i is measured from the exit pupil and D refers to the diameter of the exit pupil.

The reflecting telescope designs described later in this chapter have a circular obstruction in front of the entrance pupil. The f-number for these systems is given by[31]

$$f/\#_{Obs} = f/\# \sqrt{\frac{1}{1 - \left(\frac{D_{Obs}}{D_{Enp}}\right)^2}} \tag{7.51}$$

where $f/\#$ and $f/\#_{obs}$ are the f-numbers of the unobscured and obscured optical systems and D_{Obs} and D_{Enp} are the diameters of the obscuration and entrance pupils, respectively. Alternately, the irradiance at the detector from the obscured aperture illustrated in Figure 7.49 can be calculated directly from $\theta_{\min}$ and $\theta_{\max}$:

$$E_d = \pi L_s \left(sin^2\,\theta_{\max} - sin^2\,\theta_{\min}\right)^2 \tag{7.52}$$

This follows directly from (7.47) by integrating from $\theta_{\min}$ and $\theta_{\max}$.

We now give f-number definitions that are used to describe diffraction from a thin lens (Figure 7.50). For a thin lens, with the object an arbitrary distance

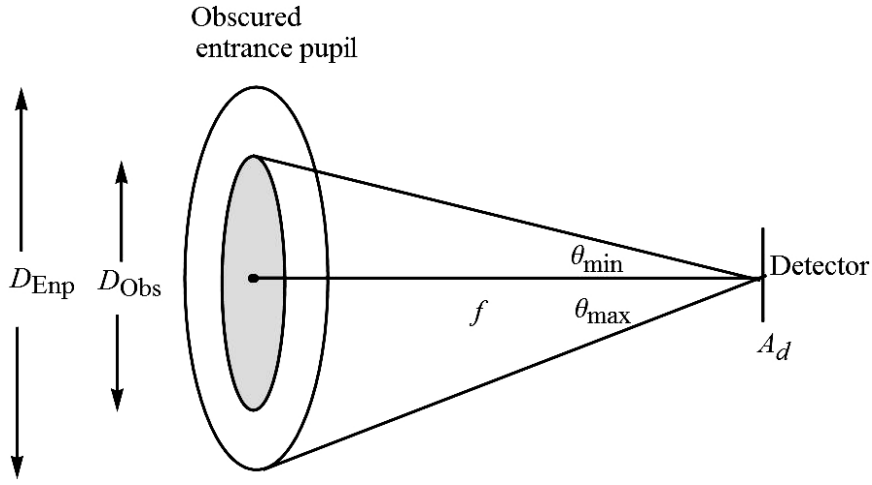


Figure 7.49 The *f*-number for an obscured aperture.

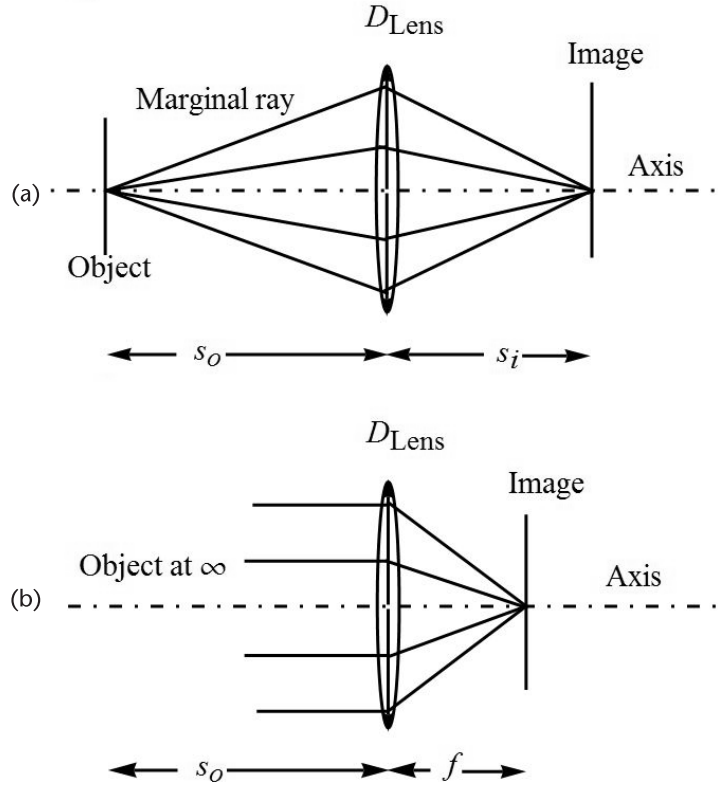


Figure 7.50 Definition of *f*-number for a thin lens in object and image space.

from the lens, the *f*-number is defined in object and image space by the following relationships[6]:

$$(f/\#)_{Object\ space,Object\ not\ at\ \infty} \equiv \frac{s_o}{D_{Lens}} \tag{7.53}$$

$$(f/\#)_{Image\ space,Object\ not\ at\ \infty} \equiv \frac{s_i}{D_{Lens}} \tag{7.54}$$

If the object is at infinity, then s_i equals the focal length of the lens and (7.54) becomes

$$(f/\#)_{Image\ space,Object\ at\ \infty} \equiv \frac{f}{D_{Lens}} \tag{7.55}$$

For a thick lens, as shown in Figure 7.51, (7.53) and (7.54) need to be generalized:

$$(f/\#)_{Object\ space,Object\ not\ at\ \infty} \equiv \frac{s_{Enp}}{D_{Enp}} \tag{7.56}$$

$$(f/\#)_{Image\ space,Object\ not\ at\ \infty} \equiv \frac{s_{Exp}}{D_{Exp}} \tag{7.57}$$

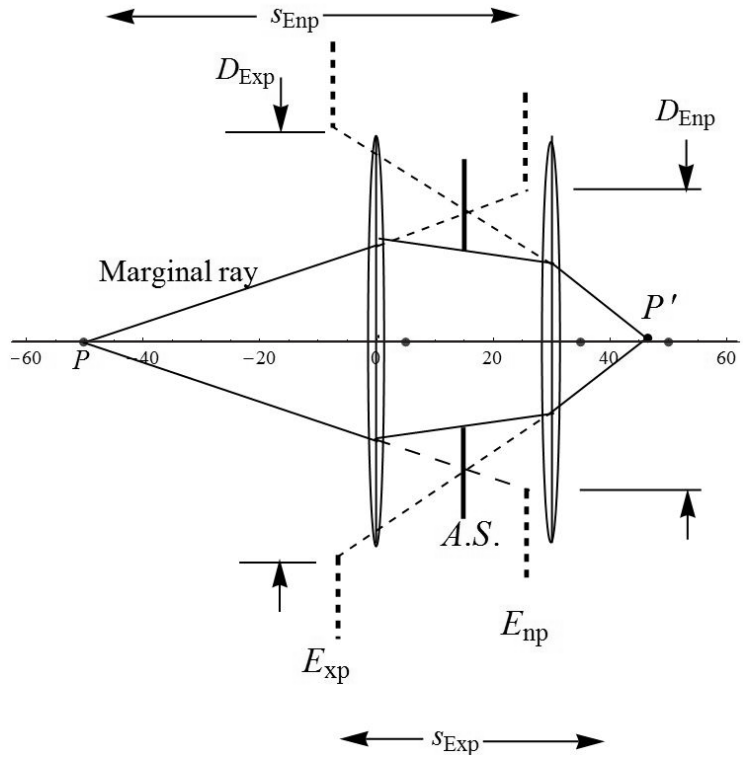


Figure 7.51 Definition of *f*-number for a thick lens in object and image space.

If the object is at infinity than (7.55) becomes[6]

$$(f/\#)_{Image\ space,Object\ at\ \infty} \equiv \frac{s_{Exp}}{D_{Exp}} = \frac{f_{eff}}{D_{Enp}} \tag{7.58}$$

In all cases the *f*-number is a measure of the solid angle subtended at the lens by either the object or the image. The definitions of (7.53) through (7.58) are useful[6] for describing the diffraction performance of a system when the object is not at infinity.

7.12 Telescopes and Angular Magnification

A distant object that subtends an angle α to the naked eye will subtend a greater angle α' when viewed through a telescope. Angular magnification is defined by

$$M_{ang} \equiv \frac{\alpha'}{\alpha} \tag{7.59}$$

The angular magnification produced by a telescope enables one to view details in a distant object that would not be seen by the naked eye and is useful in surveillance applications.

The angles α and α' are illustrated in Figure 7.52 for a Keplerian telescope. On the left, the solid lines come from a distant point on the optical axis, while the dashed lines come from a distant point below the optical axis. The first lens has a focal length of f_o and is called the *objective lens*. The second lens has a focal length of f_e and is called the *eyepiece lens*. The eyepiece lens is sometimes referred to as the *ocular*. The two lenses are separated by a distance equal to the sum of the focal lengths. Rays from point *P* on the optical axis (not shown because it is too far away) come to a focus at point *P'* in the diagram and leave the second lens traveling parallel

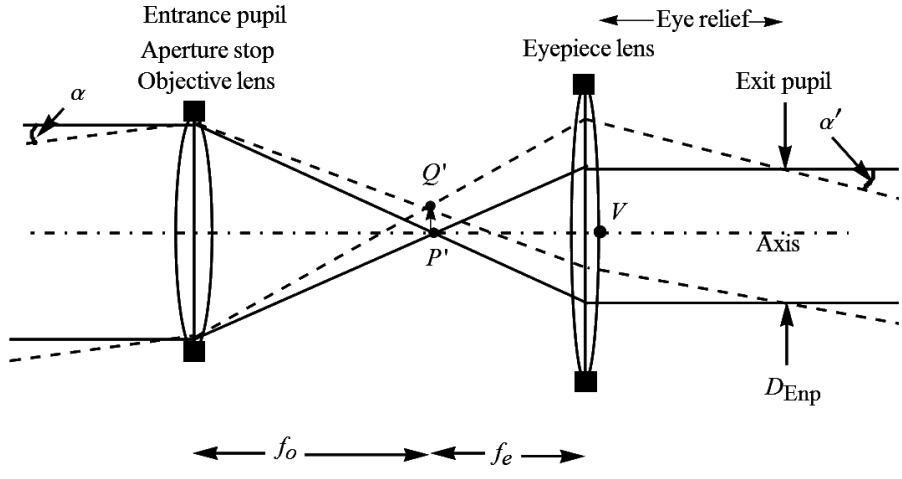


Figure 7.52 Keplerian telescope.

to the optical axis. Rays from a point on the distant object Q below the axis are brought to a focus at the point Q'. Not shown in the diagram (it is too far away) is the arrow $P\,Q$, which points down.

Because exiting rays from a distant point on the optical axis are not brought to a focus to the right of the eyepiece lens, the Keplerian telescope does not have a focal length and it is termed an *afocal optical system*.

Observe that the objective lens is the aperture stop since it is the optical element that is limiting the cone of rays from the distant on-axis point, and if the objective lens is thin, then the entrance pupil coincides with the principal plane and the image of the aperture stop formed by the objective lens is the principal plane (see Figure 7.37). The net result is that for a Keplerian telescope utilizing thin lenses, the entrance pupil and aperture stop coincide with the principal plane of the objective lens. Recall that the exit pupil is the image of the aperture stop in the eyepiece lens. Since the dashed and solid lines cross in the objective lens, the place where the dashed lines cross the solid lines to the right of the eyepiece lens is the exit pupil. An observer who positioned the entrance pupil of his eye so that it coincided with the exit pupil of the telescope would intercept all of the rays coming from the object that made it through the objective lens.

The exit pupil of the human eye varies between 2 mm in bright light to roughly 8 mm in darkness. If the telescope is meant to be used by a person, then we would want to design the system such that the exit pupil was between 2 and 8 mm, depending on lighting conditions, to match the eye pupil size of the person using the telescope.

Eye relief is another parameter that characterizes and optical system. In Figure 7.52 the point V where the optical axis intersects the eyepiece lens is called a *vertex*. The *vertex plane* is a plane perpendicular to the optical axis at a vertex. Eye relief[8,26] is the distance between the exit pupil and vertex plane associated with the point V. Recommended values for eye relief are 15 to 17 mm for typical binoculars. A greater value of 25 to 100 mm of eye relief is needed for rifle telescopic sights to avoid eye injury from rifle recoil (http://en.wikipedia.org/wiki/Eye_relief).

We will show that the angular magnification of the Keplerian telescope is

$$M_{ang} = \frac{f_o}{f_e} \tag{7.60}$$

where f_o and f_e are the focal lengths of the objective an eyepiece lens, respectively. It is usually much more practical to change the focal length of the eyepiece lens than it is to change the focal length of the objective. Typically a telescope user has a variety of eyepiece lenses and these are used to change the telescope's angular magnification.

The chief ray in Figure 7.53 comes from a point Q on a distant arrow pointing down. Since Q' is an image point, rays from this point are emerging in a variety of directions. To see the exit direction of rays from Q', construct the ray $Q'\,Q''$ through the center of the eyepiece lens. As in Figure 7.52 the angles α and α' denote the angles subtended by the target at the unaided eye and with the telescope. From Figure 7.53

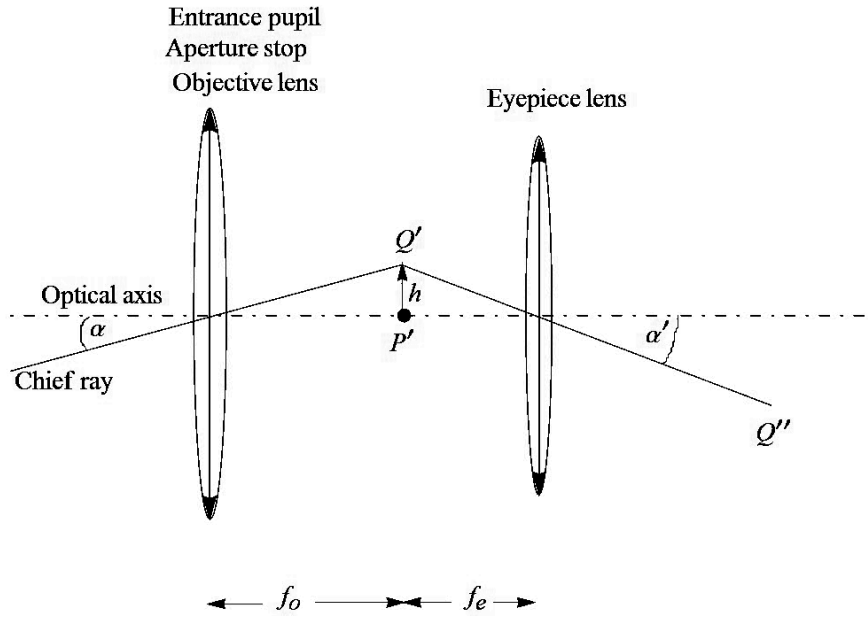


Figure 7.53 Angular magnification of Keplerian telescope.

$$\tan\alpha = \frac{h}{f_o} \text{ and } \tan\alpha' = \frac{h}{f_e} \quad \Rightarrow \quad f_o \tan\alpha = f_e \tan\alpha' \tag{7.61}$$

Typically the angle α subtended at the eye is much less than one radian and although the angle α' is larger, typically it too is small compared to one radian so (7.61) becomes

$$M_{ang} \equiv \frac{\alpha'}{\alpha} = \frac{f_o}{f_e} \tag{7.62}$$

It is instructive to calculate the FOV of the telescope shown in Figure 7.52. Let FOV_e denote the maximum field of view of the eyepiece lens and let FOV_t denote the telescope field of view. Suppose the object imaged as point P' Q' in Figure 7.53 is so large that it covers half the telescope field of view. Then

$$\alpha = \frac{FOV_t}{2} \text{ and } \alpha' = \frac{FOV_e}{2} \quad \Rightarrow \quad \frac{\alpha'}{\alpha} = \frac{FOV_e}{FOV_t} = \frac{f_o}{f_e} = M_{ang}$$

where the last two equalities follow from (7.62). Solving for FOV_t

$$FOV_t = \frac{FOV_e}{M_{ang}} = \frac{f_e}{f_o} FOV_e \tag{7.63}$$

We now find the size of the Keplerian telescope exit pupil. In Figure 7.54, triangles $A\,P'\,B$ and $C\,P'\,D$ are similar because they are both isosceles triangles and they have the same angle at point P'. Therefore,

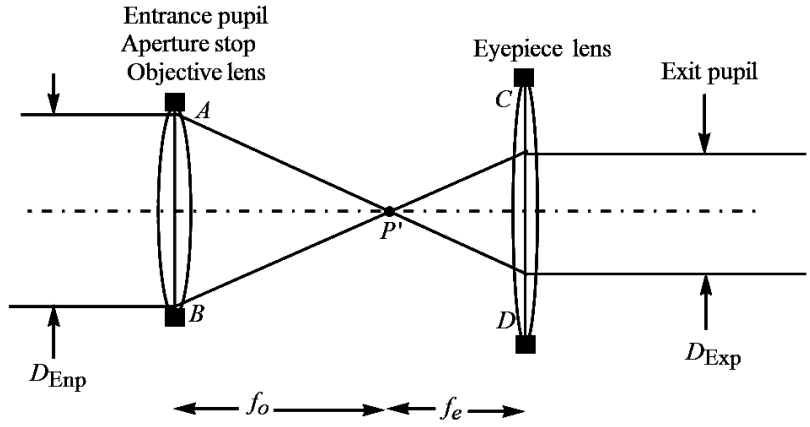


Figure 7.54 Exit pupil of Keplerian telescope.

$$\frac{D_{Enp}}{f_o} = \frac{D_{Exp}}{f_e} \quad \Rightarrow \quad D_{Exp} = D_{Enp}\frac{f_e}{f_o} = \frac{D_{Enp}}{M_{ang}} \tag{7.64}$$

When optical subsystems are joined to make an optical system, it is good design practice to match the location and size of the exit pupil of the first system with the entrance pupil of the second system. This ensures that all of the radiation output from the first system gets through to the second system. If the entrance pupil of the second system is larger than the entrance pupil of the first system and both pupils completely overlap, then all of the radiation from the first system will get through to the second system but the second system will be larger than it needs to be. Since we want the two pupils to be coincident in position along the optical axis, it is important to be able to calculate the location of the exit pupil. For that reason we calculate the location of the exit pupil for the Keplerian telescope of Figure 7.52.

Recall that the exit pupil is the image of the aperture stop in the optical elements downstream of the aperture stop. Examining Figure 7.54, the only element downstream of the aperture stop is the eyepiece lens. Relative to the eyepiece lens, the object distance $s_o = f_o + f_e$. Denote the distance of the exit pupil to the right of the principal plane of the eyepiece lens by s_{Exp}. Then using Gauss's thin lens equation,

$$\frac{1}{f_o + f_e} + \frac{1}{s_{Exp}} = \frac{1}{f_e} \quad \Rightarrow \quad s_{Exp} = \frac{f_e\left(f_o + f_e\right)}{f_o} \tag{7.65}$$

In Figure 7.55 the principal plane of a camera lens with focal length f_c is placed at the exit pupil of the telescope. To more clearly illustrate the effective focal length of the system, here the entrance pupil of the camera is drawn significantly larger than the exit pupil of the telescope. In practice, the entrance pupil of the camera would be the same size or a little larger than the telescope exit pupil. The distant point P, which is on the optical axis, is focused at point P'' in the camera. To find the effective focal length of the optical system, we forward trace the top marginal ray and back trace the marginal rays from P''. The two rays meet at point R. Pass a plane through R perpendicular to the optical axis. This is the second principal plane of the

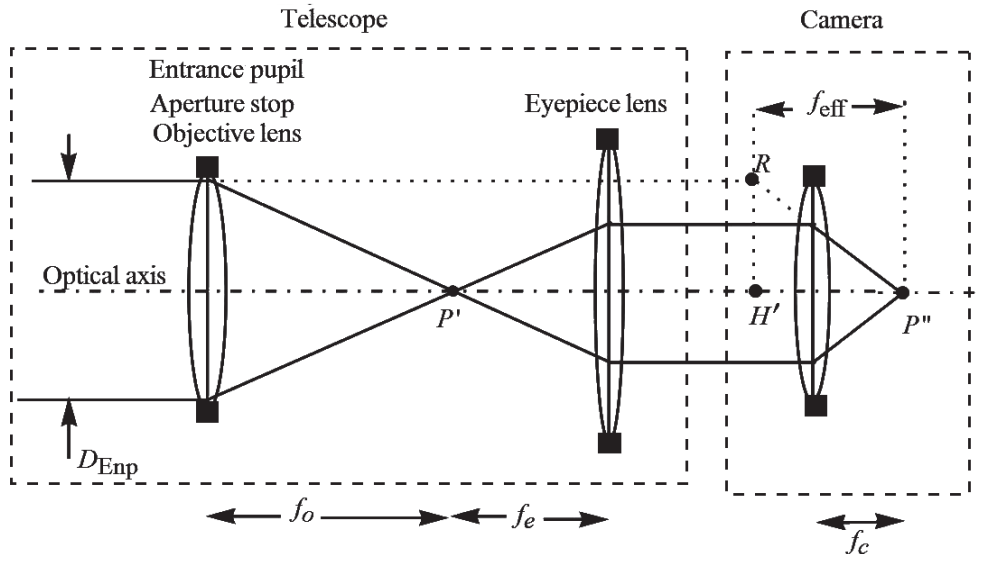


Figure 7.55 The effective focal length and entrance pupil of a telescope-camera optical system.

optical system and it intersects the optical axis at point H'. The system focal length, here called the *effective focal length*, is the distance f_{eff} illustrated in Figure 7.55.

We wish to show that

$$f_{eff} = \frac{f_o f_c}{f_e} \tag{7.66}$$

Taking the tangent of the two half angles at point P' in Figure 7.56,

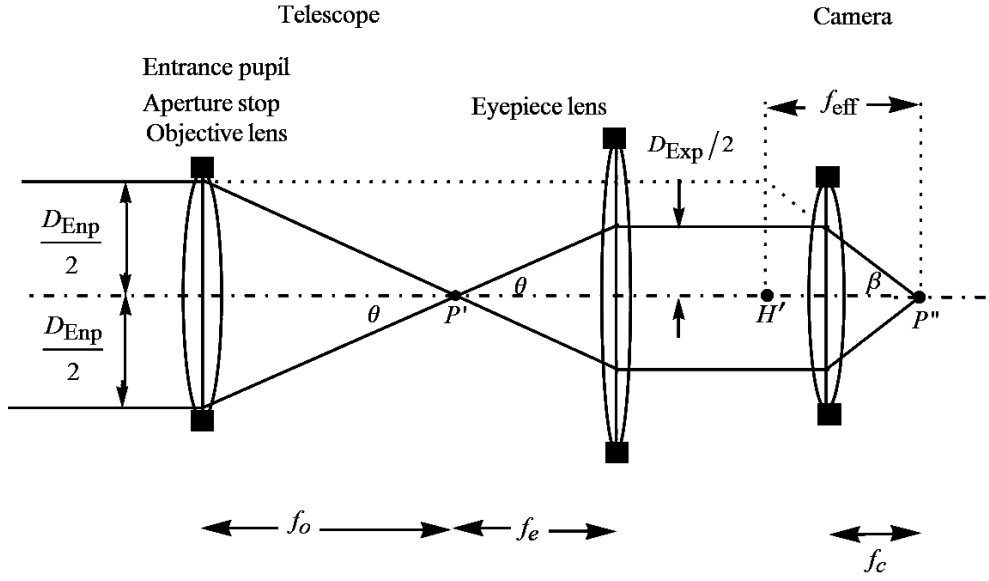


Figure 7.56 Calculating effective focal and f/# length and f/# of a telescope-camera system.

$$\tan\theta = \frac{\frac{D_{Enp}}{2}}{f_o} = \frac{\frac{D_{Exp}}{2}}{f_e} \quad \Rightarrow \quad \frac{D_{Enp}}{D_{Exp}} = \frac{f_o}{f_e} \tag{7.67}$$

in agreement with (7.64). There are two triangles with half angle β at the vertex P'' and this enables the tangent of β to be expressed in two different ways:

$$\tan\beta = \frac{\frac{D_{Enp}}{2}}{f_{eff}} = \frac{\frac{D_{Exp}}{2}}{f_c} \quad \Rightarrow \quad f_{eff} = \frac{D_{Enp}}{D_{Exp}} f_c \tag{7.68}$$

The result (7.66) is readily obtained by combining (7.67) and (7.68).

The *f*-number of the telescope-camera system is now readily computed. Let D_{obj} denote the diameter of the objective lens, which is equal to the diameter of the entrance pupil. Then

$$f/\# \equiv \frac{f_{eff}}{D_{Enp}} = \frac{f_o f_c}{f_e D_{obj}} \tag{7.69}$$

where the last equality follows from (7.66).

From (7.62) it would seem that one could obtain an arbitrarily large angular magnification by making f_e arbitrarily small and making f_o arbitrarily large. The angular magnification does get larger with increasing f_o and decreasing f_e but notice from (7.69) that this implies the *f*-number also gets larger so that as one tries to get more angular magnification, the image gets dimmer. As indicated by (7.66) and (7.69), the cost of increasing angular magnification is a dimmer image and a smaller *FOV*. One cannot resolve detail [see (7.24)] that has an angular separation of less than $1.22\lambda/D_{enp}$. Attempting to seek detail with an angular separation of less than $1.22\ \lambda/D_{enp}$ is paid for with a dimmer image, a smaller *FOV*, and, because of diffraction limits, is of no value. This state of affairs is described by the phrase *empty magnification*. Empty magnification[22] is magnification above the level of maximum useful magnification and does not contribute useful resolving power. In a visual telescope, empty magnification is angular magnification in excess of $20\ D_{enp}$ where D_{enp} is the diameter of the telescope entrance pupil in inches.[22]

The field of view of the Keplerian telescope given by (7.63) was determined by the properties of the eyepiece lens. If we wanted to further limit the telescope FOV, a field stop could be added as shown in Figure 7.57. Let D_F denote the diameter of the field stop and FOV_t denote the field of view of the telescope. The two chief rays shown in the figure come from extreme points above and beneath the optical axis from the distant object that just make it through the telescope. Examining the triangle formed by the chief rays at the principal plane of the objective lens, we get

$$\tan\frac{FOV_t}{2} = \frac{\frac{D_F}{2}}{f_o} \tag{7.70a}$$

Typically the *FOV* of a telescope is small so the tangent of the angle can be replaced by the following angle:

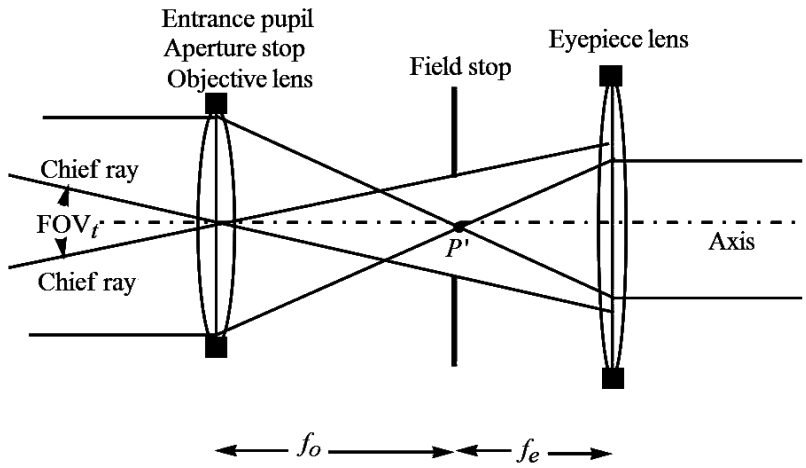


Figure 7.57 A field stop has been added to the Keplerian telescope.

$$FOV_t = \frac{D_F}{f_o} \tag{7.70b}$$

The actual value for the telescope field of view is the smaller of (7.70a) and (7.63).

A *reticle* is an optical element[22] located in an image plane, containing a pattern that assists in pointing an instrument or measuring target characteristics. It may be two crossed lines or a more complex pattern. Suppose two thin wires were attached to the field stop in Figure 7.57 so that they were in the plane of the field stop and intersected at a right angle in the center of the field stop.

Then when viewed through the eyepiece lens the field stop would appear as a circular aperture with two cross hairs intersecting at right angles.

In Figure 7.57 the objective and eyepiece lenses are separated by the sum of their focal lengths and this results in parallel rays exiting the telescope. How does one see an image if the rays emerging from the telescope are parallel to the optical axis? Parallel rays correspond to rays from a distant object and the normal naked eye can bring these rays to a focus on the retina. So when looking through the telescope of Figure 7.57, where the objective and eyepiece lens are separated by the sum of their focal lengths, it appears to the observer's eye that the object is at infinity and yet the object subtends a finite angle at the observer's eye.

Suppose the separation between the objective and eyepiece lens is less than $f_o + f_e$. Then the rays emerging from point P' would have greater curvature at the eyepiece lens with the result that the rays exiting the eyepiece lens now diverge. Back tracing these diverging rays, the object appears to come from the point from which these rays diverge. If the divergence of the rays exiting the eyepiece lens is not too severe, the eye accommodates the diverging rays and the object can still be seen clearly. This shows that it is not critical for the objective and eyepiece lens to be separated by exactly the sum of their focal lengths.

Suppose the separation between the eyepiece and objective lens is greater than $f_o + f_e$. Then the rays emerging from point P' would have reduced curvature at the eyepiece with the result that the eyepiece lens would bring the rays to a focus to the

right of this lens and by putting film or a detector array at this location an image of the distant object could be made.

When viewed through the eyepiece, the Keplerian telescope illustrated in Figure 7.52 produces an inverted image. Examining Figure 7.52, $P'\,Q'$ is erect indicating the object $P\,Q$ (not shown) points down. If the eye is placed at the exit pupil, the ray from the tip of the down-pointing arrow will point down. The mind perceives the upside down arrow on the retina as being right side up in disagreement with the actual arrow $P\,Q$, which points down. *Terrestrial telescopes* produce an erect image.[22] Because the telescope illustrated in Figure 7.52 produces an inverted image, it is sometimes referred to as an *astronomical telescope*. Add an erecting lens as shown in Figure 7.58 to convert the Keplerian telescope into a terrestrial telescope.

It is easy to see that the telescope of Figure 7.58 will produce an upright image. The arrow $P'\,Q'$ is upside down indicating the arrow $P\,Q$ (not shown because it is too far away) is right side up. The erecting lens produces the image $P''\,Q''$, which is erect. The ray $Q''\,Q'''$ is traveling down, indicating that the top of the arrow will be pointing down on the retina. However, the arrow will be perceived as pointing up in the same direction as the arrow $P\,Q$.

The index of refraction of glass depends on the wavelength of light and because of this red light and blue light from the objective come to a focus at different points on the optical axis. The net result is that for the telescope illustrated in Figure 7.52, red and blue light from a point on the source come to a focus on different parts of the retina or in the focal plane of the camera. This increases the psf and leads to a less distinct image. An *achromatic lens* is one consisting of two or more elements, usually crown and flint glass that have been designed so that two extreme wavelengths (typically red and blue) come to a focus at a single point on the optical axis. The increase in the psf caused by a refractive index that varies with wavelength is a manifestation of *chromatic aberration*. One way to reduce chromatic aberration is to use achromatic lenses. Because the law of reflection [see (7.4)] is wavelength independent, another way to reduce chromatic aberration is to use mirrors where possible.

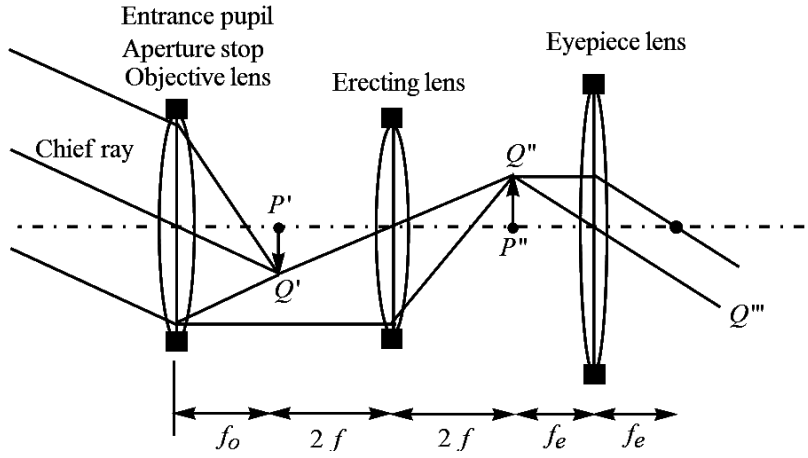


Figure 7.58 Terrestrial telescope.

Figure 7.59 shows three reflecting telescope designs. The primary mirror in the Newtonian telescope is a paraboloidal surface. The secondary mirror is plane and makes a 45-deg angle with the optical axis and is placed so that the focal point is just outside the telescope tube. This design eliminates spherical aberration at the prime focus. The Cassegrain configuration has several designs. In the classical Cassegrain design, the primary is paraboloidal, the secondary is a hyperboloid and the focal points of the parabolical and hyperbolical mirrors coincide. The parabolical mirror has the property that it focuses all rays parallel to the optical axis to a single point and the convex hyperbolical mirror has the property that all rays incident on one

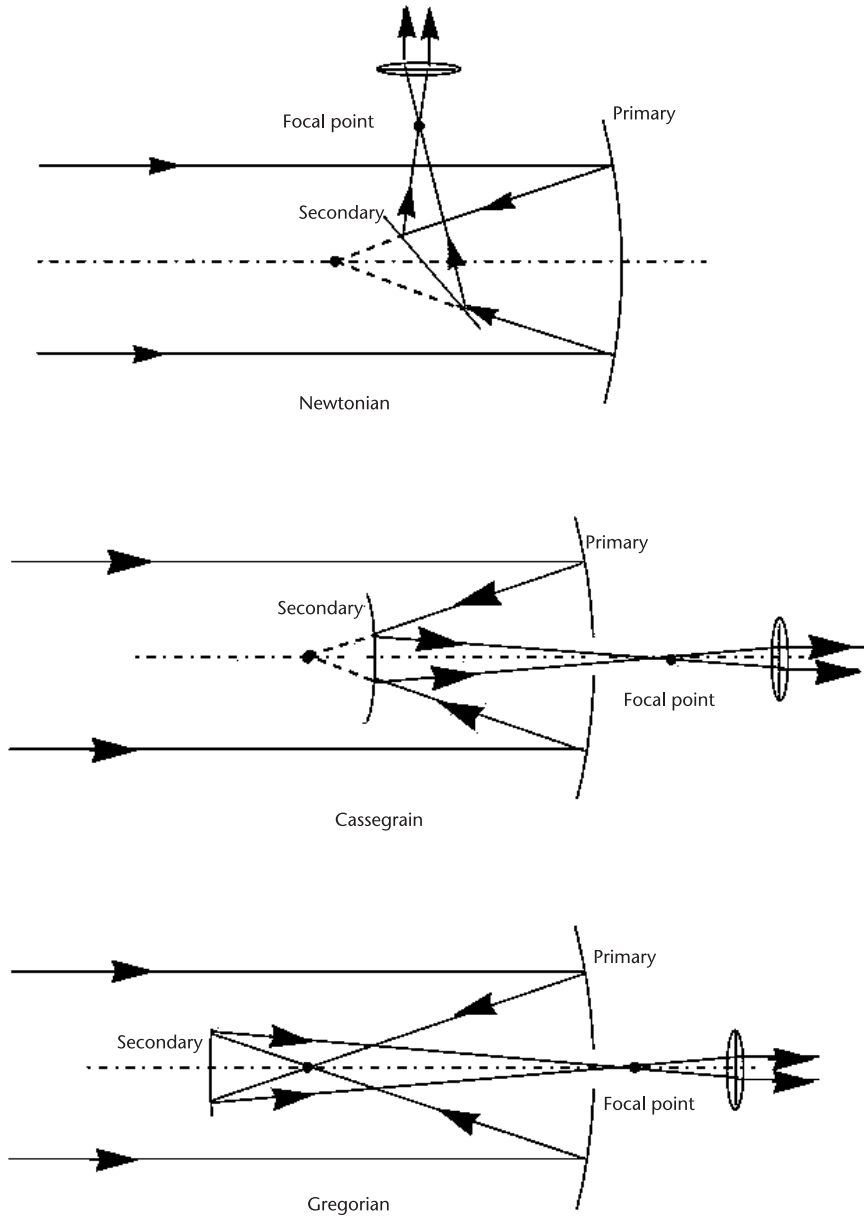


Figure 7.59 Newtonian, Cassegrain, and Gregorian reflecting telescopes.

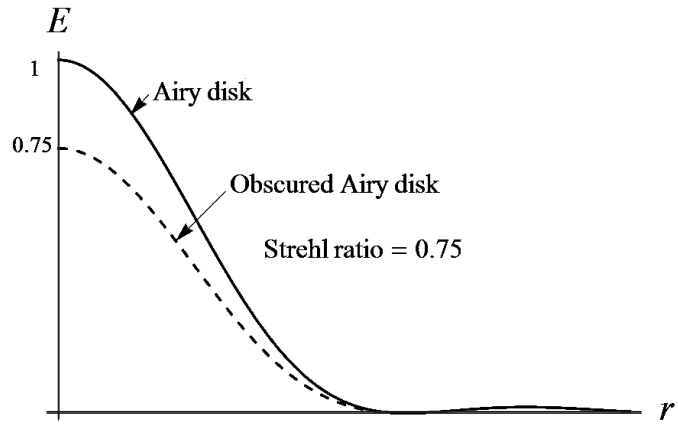


Figure 7.60 Definition of Strehl ratio.

focus of the hyperboloid get focused to the second focus of the hyperboloid.[10] The classical Cassegrain design eliminates spherical aberration. In the Ritchey-Chrétien Cassegrain design,[22] the primary is a concave hyperboloid and the secondary is a convex hyperboloid. In the Dall Kirkham Cassegrain design, the primary is a prolate ellipsoid and the secondary mirror is spherical.[22]

The reader may wonder if the central part of the field of view is obscured by the central mirror in these designs. For sufficiently distant objects, the central part of the FOV is not obscured by the secondary mirror and the reason for this is that distant on-axis points are seen by the primary mirror. From Figure 7.59 we observe that for distant objects the incident rays are parallel to the optical axis and so are reflected from the primary mirror and then after being reflected by the secondary mirror to a focal point they pass through the eyepiece lens and enter the eye as collimated light. The central obscuration of the designs shown in Figure 7.59 does reduce the radiation reaching the sensor and for this reason does effectively increase the *f*-number of the optical system as indicated by (7.51). Another effect of the central obscuration is that it takes energy from the central maximum of the diffraction psf and distributes it into the rings of the diffraction pattern. This effect is quantized by a single number termed the *Strehl ratio* (Figure 7.60). The *Strehl ratio* is defined[5] as the irradiance at the center of the optical systems impulse response, divided by that at the center of a diffraction-limited impulse response from a clear circular aperture. The maximum value of the Strehl ratio is 1. Aberrations reduce the Strehl ratio as does the central obscuration present in the reflecting telescopes illustrated in Figure 7.59.

7.13 Modulation Transfer Function

If the object plane is sufficiently far away, the thin lens system shown provides an image one focal length from the principal plane, which also coincides with the entrance pupil. In typical imaging systems, the collection aperture is located at the limiting pupil function $p(x,y)$. We know from Chapter 4 that the system shown has both a coherent transfer function and an incoherent transfer function (or optical

transfer function, *OTF*). While we are not usually interested in the coherent transfer function, it is easier to start with it to develop the incoherent transfer function. For the rectangular aperture shown, the coherent transfer function is [see (4.52c)]

$$H_{Coh}(\xi,\eta) = rect\left(\frac{x}{L_x},\frac{y}{L_y}\right)_{x=\xi\lambda f, y=\eta\lambda f} \tag{7.71}$$

Recall that the Fourier transform of this coherent transfer function gives the coherent impulse response (i.e., a two-dimensional *sinc* function). A large aperture corresponds to a small impulse response and consequently higher spatial resolution in the image plane. Also, from Chapter 2, the *rect* function transitions from 1 to 0 at arguments of 1/2. The transition occurs when $\xi = L_x/(2\lambda f)$ and $\eta = L_y/(2\lambda f)$. These are the cutoff spatial frequencies in the image plane. Note that the units are cycles per meter or cycles per millimeter depending on the units of length used for aperture size, wavelength, and focal length.

To determine the OTF of the system shown in Figure 7.61, the normalized autocorrelation of the coherent transfer function must be found [see (4.49)}:

$$OTF_{InCoh}(\xi,\eta) = H_{InCoh}(\xi,\eta) = \frac{\gamma_{H_{Coh}}(\xi,\eta)}{\left|\gamma_{H_{Coh}}(0,0)\right|} \tag{7.72}$$

Although (7.72) appears complicated, $OTF_{InCoh}(\xi, \eta)$ is simply the autocorrelation of the coherent optical transfer function with the peak value (at the function origin) set to 1. With the substitutions given in (7.71) the system $OTF_{InCoh}(\xi, \eta)$ is

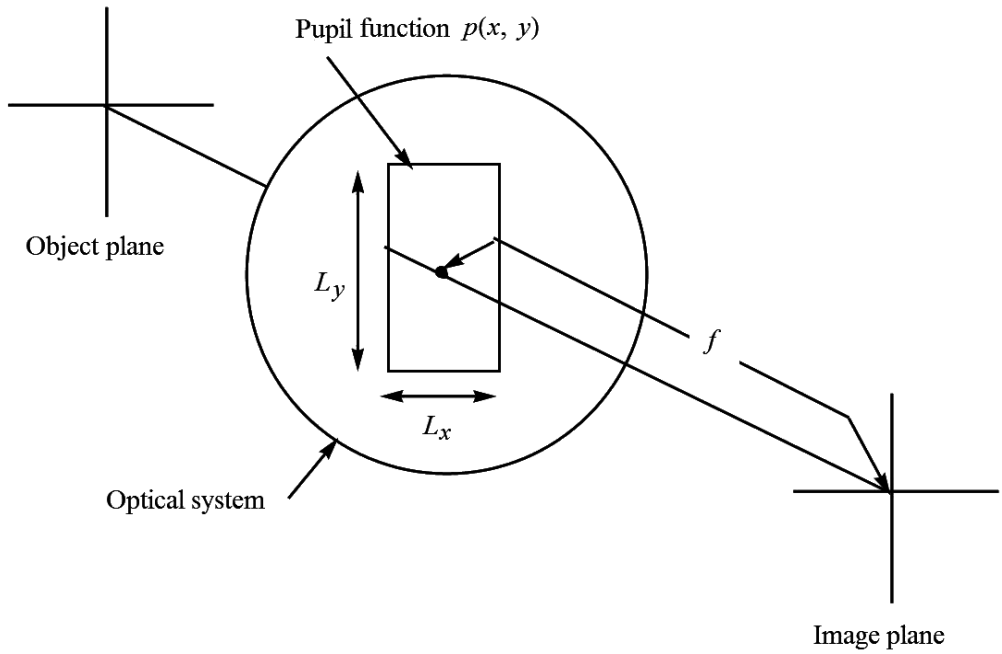


Figure 7.61 Optical system with pupil function.

$$OTF_{InCoh}(\xi, \eta) = tri\left(\frac{\xi \lambda f}{L_x}, \frac{\eta \lambda f}{L_y}\right) \tag{7.73}$$

The cutoff spatial frequencies for OTF_{InCoh} occur at $\xi_{cut} = L_x/(\lambda f)$, $\eta_{cut} = L_y/(\lambda f)$, which are twice as large as the coherent case.

Three very important characteristics of the transfer functions must be discussed here. First, the functions describe the transfer of spatial frequencies onto the image plane where they are defined in cycles per distance. It is not always convenient to represent spatial frequencies in these units, so they can be converted to angular spatial frequencies simply by dropping the f in (7.71) and (7.73). Although cycles per distance units are convenient for image plane calculations, many times we are interested in the angular spatial frequency response of the system. For example, many sensor performance specifications are given in units of cycles per milliradian.

The second characteristic is the segmentation of the *OFT* into the modulation transfer function (*MTF*) and the phase transfer function (*PTF*). The *MTF* is the magnitude or modulus of the OTF and the PTF is the phase of the *OFT*. In some cases, the *OTF* gives a negative value for some range of the spatial frequencies. When this occurs, the *PTF* provides the phase shift necessary to change bright regions in the image to dark, and dark regions to bright (i.e., phase reversal). This occurs for the spatial frequencies where *PTF* is negative. This is not a desirable design characteristic, so it is not the intent of sensor designers to have their systems operate in this mode. Severe defocus is one of the primary causes of phase reversal.

Finally, the third important characteristic is that of signal transfer. Consider a slice through the *OTF* in the horizontal direction for both the coherent and incoherent optical transfer functions (Figure 7.62). For the coherent transfer function, radiation up to the cutoff frequency transfers undegraded. This is not the case for the incoherent transfer function where the amplitude of the higher frequencies is reduced. The amplitude of an output signal with a frequency of half the cutoff will be reduced by a one-half. This transfer is with respect to the zero frequency signal because there is some transmission loss even for the dc (zero frequency) signal

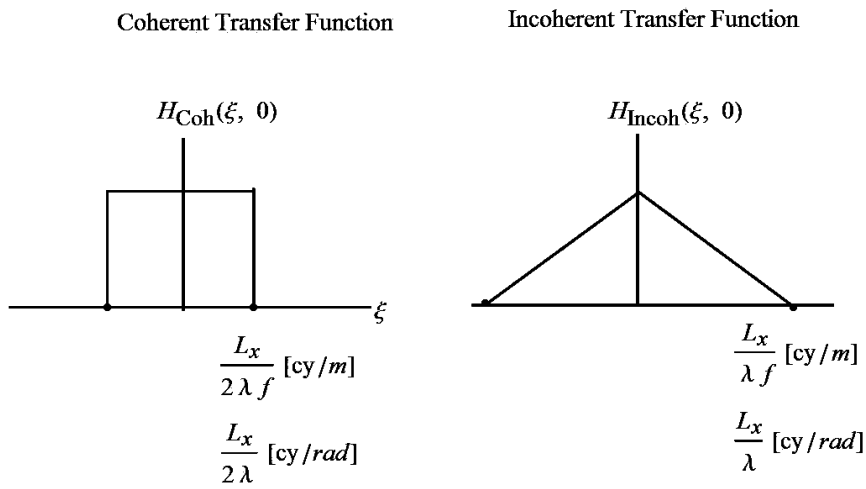


Figure 7.62 Coherent and incoherent transfer functions.

through the optical system. In summary, the resolution of the coherent system is less than that of the incoherent system, but signals of all frequencies less than the cutoff transfer through the system in the same manner. The resolution of the incoherent system is twice that of the coherent system, but higher spatial frequency signals suffer more degradation as they pass through the system. In either system, no signals higher than the cutoff frequencies are seen in the output image.

For simplicity, the rectangular aperture was selected to illustrate coherent and incoherent transfer function concepts. We now apply what we have done for the rectangular aperture to the more common circular aperture. An unobstructed circular aperture placed in the system of Figure 7.58 yields a coherent transfer function:

$$H_{Cob}(\rho) = cyl\left(\frac{r}{D}\right)_{r=\rho\lambda f} \tag{7.74}$$

where ρ is the circularly symmetric spatial frequency. The incoherent *OTF* is the normalized autocorrelation of this function:

$$OTF_{InCoh} = \frac{\gamma_{H_{Cob}}(\rho)}{\left|\gamma_{H_{Cob}}(0)\right|} \tag{7.75}$$

and the incoherent MTF is the magnitude of this function. While a closed-form solution for the autocorrelation of the cylinder function is a straightforward geometry problem, the derivation is cumbersome. We know that the function is twice the width of the original cylinder function, so the cutoff frequency is twice that of the coherent system. Also, we know that the correlation degrades the higher frequencies because of the smoothing operation. The evaluation of the autocorrelation function of (7.75) is provided by Gaskill[7] as

$$MTF(\rho) = \frac{2}{\pi}\left[cos^{-1}\left(\frac{\rho}{\rho_c}\right) - \frac{\rho}{\rho_c}\sqrt{1-\left(\frac{\rho}{\rho_c}\right)^2}\right] \tag{7.76}$$

where $\rho_c = D/(\lambda f)$ in cycles per unit length (usually, millimeters or meters) or $\rho_c = D/\lambda$ in cycles per radian. Here D is the diameter of the circular entrance pupil. Equation (7.76) is valid only for spatial frequencies less than the cutoff frequency, and the MTF is zero at spatial frequencies above the cutoff frequency. A comparison of this function with the MTF of a rectangular aperture of the same size is shown in Figure 7.63.

The Fourier transform of either function shown in Figure 7.63 yields the psf or blur spot of the corresponding diffraction-limited optical system. The transfer function for optics with aberrations must be treated separately.

Example 7.4

Determine the MTF, plotted with units of cycles per milliradian and cycles per millimeter, for a diffraction-limited optical system in the longwave region. Assume that the entrance pupil has a diameter of 20 cm and a focal length of 60 cm. Also, plot the psf in milliradians and in millimeters. We obtain a solution with an assumed

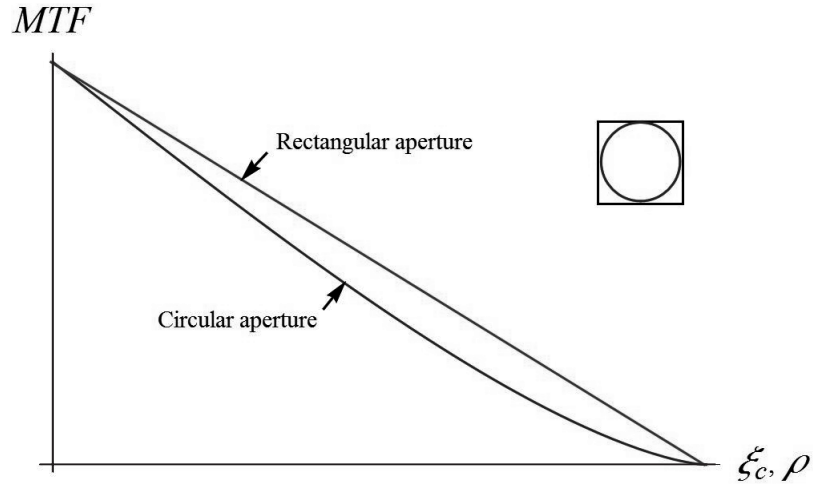


Figure 7.63 MTFs for rectangular and circular apertures.

wavelength of 10 μm, which gives an angular spatial cutoff frequency of 20 cycle/mrad and the focal plane spatial frequency is 33.3 cycle/mm. We use (7.76) for the MTF of a sensor with a circular entrance pupil and (7.22) for the psf. The MTF is first plotted (Figure 7.64) in angular spatial frequency and then in distance spatial frequency. The corresponding psfs are also shown. Recall that they are the Fourier transform of the transfer functions. All plots shown are circularly symmetrical.

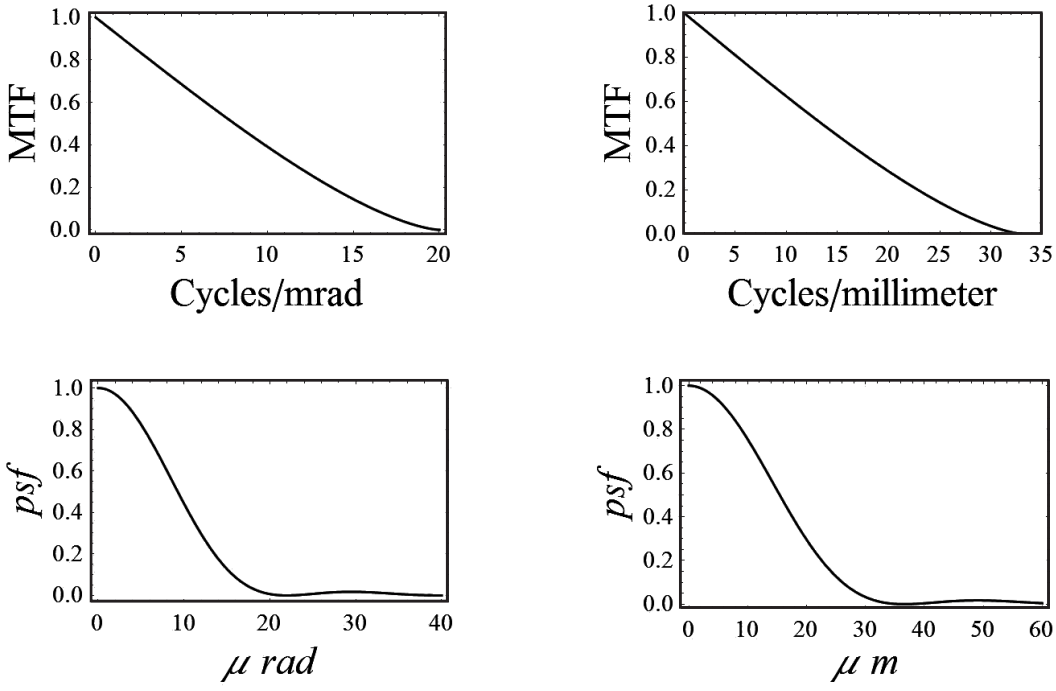


Figure 7.64 MTF and psf results.

7.14 Aberrations

The first-order imaging system analytical techniques given in the earlier sections were only approximations of actual optical system behavior. The differences in the actual behavior and the first-order estimates are known[10] as *aberrations*. Hecht and Zajac[10] provide two main classifications of aberrations: chromatic aberrations and monochromatic aberrations. Chromatic aberrations occur because the index of refraction for a particular material is a function of frequency (color in the visible spectrum). It can be shown using the lensmaker's formula (7.6) that the focal length of a lens and its corresponding image plane location change with a change in the lens refractive index. The effect of chromatic aberration within an optical system can be determined by performing a number of imaging calculations over a band of refractive indices across the waveband. A number of computer ray-tracing programs can easily perform this task.

Aberrations can occur even when the light is monochromatic. They can degrade the clarity of the image and can deform the image. Spherical aberration, coma, astigmatism, Petzval field curvature, and distortion are all monochromatic aberrations.

Spherical aberrations evolve from the first-order approximations that we made of surface power. In reality, the refractive action at a spherical surface has higher order terms:

$$\frac{1}{f} = \frac{n_2 - n_1}{R} + h^2 \left[\frac{n_1}{2o} \left(\frac{1}{o} + \frac{1}{R^2} \right)^2 + \frac{n_2}{2i} \left(\frac{1}{R} - \frac{1}{i} \right)^2 \right]$$

The higher order terms shown here would give a third-order approximation. Note that the additional term is a function of ray height above the optical axis. That is, the error between the first-order and third-order approximations is a function of the ray height. A large h causes the light to focus at a shorter distance from the surface as shown in Figure 7.65.

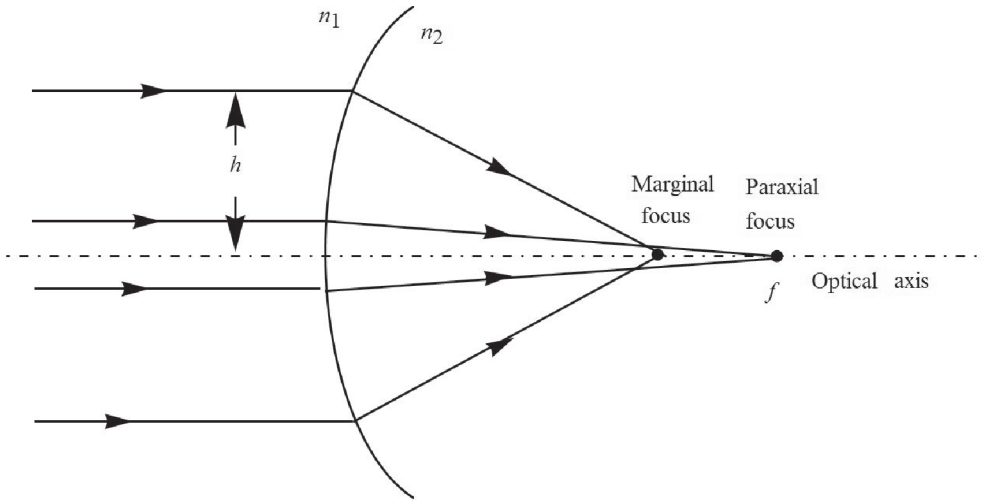


Figure 7.65 Spherical aberration.

The paraxial focus position is valid (Figure 7.65) for those rays that enter the surface near the optical axis. For a given application that uses spherical lenses, h is limited and a "circle of least confusion" is found where the image plane is set between the marginal focus and the paraxial focus. This position provides an image with a minimum diameter. In some cases, where large apertures are required (small f-number), surfaces and lenses are constructed with aspheric surfaces that correct for spherical and other aberrations. In lens systems, the resultant spherical aberration is a summation of the contribution from each surface, so the overall spherical aberration can sometimes be corrected with a series of spherical surfaces. Also, optical designers can minimize spherical and other aberrations by adding more lenses of lower power to the system.

Spherical aberrations occur on the optical axis, or on-axis. An off axis aberration is *coma* that occurs for off-axis image points. The term *coma* comes from a Greek word meaning "tail of a comet." The imaging planes or principal planes within a lens are actually curved, giving a transverse magnification that differs for off-axis rays. In short, coma is the result of an error in off-axis magnification. The resulting image of a point appears to be a comet-like image (a round spot with a trailing tail).

Astigmatism is another image-degrading aberration. It occurs when an image point lies an appreciable distance from the optical axis such that rays strike the lens in an asymmetric manner. Recall that the plane defined by the chief ray and the optical axis is called the *meridional plane* and the plane normal to the meridional plane at the optical axis is called the *sagittal plane*. Astigmatism occurs when the image location in the meridional plane does not occur at the same position as that of the sagittal plane. For example, if the image of a point in the sagittal plane is a vertical straight line, the image at the meridional focus will be[10] a horizontal line. The best image is between these two line images. Astigmatism also occurs when the curvature of the lens is different in one direction versus another. Lens astigmatism is purposely used to correct vision from a distorted eye in prescription lenses.

Two other aberrations are *field curvature* and *distortion*. Field curvature, or Petzval field curvature, results from too many positive lenses. The image surface is curved or spherical in shape, so images projected onto flat surfaces appear sharp near the optical axis and blurred on the edges. Field curvature can be flattened by balancing the number and power of the positive and negative lenses in the system. Distortion occurs when the magnification varies as a function of radial image location. Therefore, the image size varies as a function of distance from the optical axis. The two types of distortion are pincushion and barrel as shown in Figure 7.66.

Holst[11] points out that it is convenient to represent aberrations as a single transfer function where the overall optical MTF can usually be represented as the product of a diffraction and an aberration MTF:

$$MTF = MTF_{Diff}\, MTF_{Aberr} \tag{7.77}$$

Aberrations can be approximated by a Gaussian function

$$MTF_{Aberr}(\rho) = Gaus\left(\sqrt{2\pi}\,\sigma\,\rho\right) \tag{7.78}$$

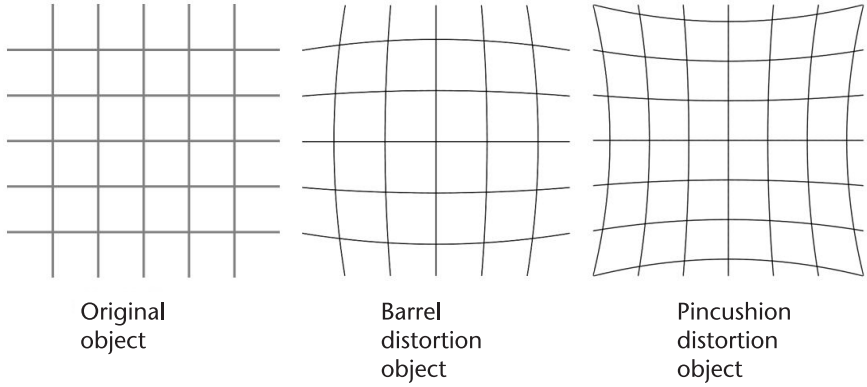


Figure 7.66 Distortion.

where σ is a parameter that describes the magnitude of the aberration and ρ is a spatial frequency in the radial direction. The Night Vision Integrated Performance Model (NV-IPM), uses this approach. Holst[11] points out that σ is not directly measureable because any optical measurement includes both the diffraction and the aberration effects. It can be determined given a measurement with the calculation of the diffraction effects.

7.15 Optical Materials

The selection of lens materials involves the consideration of many issues. A primary consideration is that of optical transmission. Many materials that transmit visible light are opaque in the infrared wavelengths. Examples are normal window glass, plexiglass, and quartz. Germanium is transmissive from 2 μm to well past 15 μm but is is opaque over EO wavelengths (0.4 to 2.0 μm). Zinc selenide is a broadband glass that transmits EO, midwave, and longwave light. Some families of glasses have a wide range of transmissions. An example is the class of IRTRAN glasses that are tailored for use in infrared systems. EO system lenses are constructed of fused silica, BK7, Zerodur, barium fluoride, and many other types of glass.

The transmission of system lenses is usually grouped into a single system optical transmission $\tau_{optics}(\lambda)$. The transmission of an optical system affects the amount of flux that propagates through the system and falls on the detector. A large transmission in the band of interest is desired. Also, the transmission is usually a strong function of wavelength, so the spectral transmission must be known along with the source spectrum and atmospheric transmission to determine the flux on the detector. The detector response is also a function of wavelength. All of these spectral quantities are integrated over wavelength to give accurate radiometric quantities.

Lens selection must also include index of refraction considerations. If the index varies as a strong function of wavelength, large chromatic aberrations occur. Sometimes a lens is constructed from two different materials with spectral index functions (dispersions) that cancel each other. Such lenses are *achromats*. Typical indices of refraction are around 1.5 for EO lens materials. Infrared lens materials

usually have a higher index of refraction. Zinc selenide's index varies (depending on wavelength) from 2.4 to 2.7. Germanium has an index of refraction of around 4 in the midwave.

Other material considerations are thermal properties, elastic constants, hardness, and other mechanical properties. Lens design and material selection are extremely complicated processes. Usually, a system-level engineer or scientist does not perform this level of development, because there are other optical professionals that spend their lives producing optical designs. These professionals are usually consulted during the development of an imaging infrared or EO system.

7.16 Cold Stop and Cold Shield

The housing of an infrared camera is typically close to ambient temperature. At ambient temperature many photons are emitted by the camera housing and these photons compete with photons emitted by the target and the background. The photons emitted by the camera provide an extraneous signal that masks the target and background signals. The SNR would be improved if the noise photons from the housing could be reduced and this is done by a *cold stop*. As stated by Dereniak and Boreman[6]:

> The detector is housed in a vacuum enclosure called a *dewar* which is usually the only part of the system that is cooled. Within the dewar a cooled aperture called a *cold shield* is placed adjacent to the detector plane to limit the angle over which the detectors receive radiation. The cold shield passes radiation from scene sources and blocks radiation from internal sources. The degree to which the cold shield accomplishes this task is quantified by a parameter called the *cold-stop efficiency*. If the cold shield is placed at a plane that is an image of the aperture stop, then the detectors will receive flux only from scene sources (100% cold-stop efficiency, assuming no scattered radiation), and the cold shield is called a *cold stop*.

A cold stop is illustrated in Figure 7.67. If the cold stop was made larger than the exit pupil or moved so that it was no longer at the exit pupil, then it could have less than 100% cold-stop efficiency and would be referred to as a cold shield. One way to get a dual field of view imager is to change the focal length but keep the size of the entrance pupil constant. This implies that the pencil of rays converging at P in Figure 7.67 is smaller for the narrower FOV system. The result is that a single cold shield cannot be a cold stop for both fields of view.

7.17 A Typical Optical System

This chapter would not be complete without a description of a typical optical system. There are too many different types of optical systems to begin to address their individual characteristics. These include microscopes, telescopes, night sights, target acquisition sensors, navigation sensors, cameras, camcorders, and so forth. Here, we describe one[27,28] state-of-the-art third-generation FLIR.

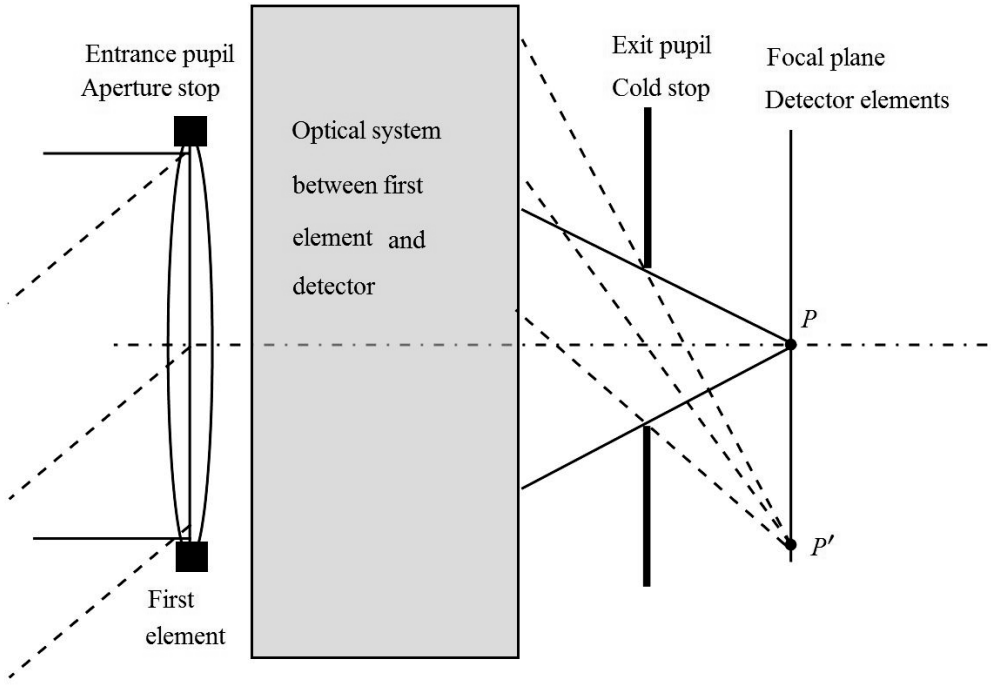


Figure 7.67 Cold stop.

As defined[28] by NVESD, a third-generation (3G) FLIR system includes the use of large-format, dual-band focal planes in a dual or variable *f*-number optical system. A 3G demonstration unit has been fabricated with the properties given in Table 7.1.This table indicates that the 3G FLIR demonstrator has a ultra-narrow FOV (UFOV), a narrow FOV (NFOV), a wide FOV (WFOV), and an extra-wide FOV (XFOV). The UFOV is 0.9 × 1.2 deg and the XFOV is 10.4 × 13.8 deg. Abbreviations are as follows: effective focal length (EFL), entrance pupil diameter (EPD), and focal plane array (FPA). An optical schematic is shown in Figure 7.68.

To understand how this optical system works look at the XFOV system in Figure 7.68. Radiation comes in and is successively reflected off mirrors 1, 2, and 3 and then, after passing through optical element 4, is focused on the FPA. Optical element 4 is actually two lenses. To switch from the XFOV to the WFOV, one of the lens elements in optical element 4 is moved axially to effectively become optical element 5 in the WFOV configuration. To go from the WFOV to the UFOV, mirror 6 is moved out of the way, allowing the radiation to successively strike mirrors 7, 8, 9, 10, and 11. The radiation then follows the same optical path as for the WFOV to get to

Table 7.1 Parameters for the Third-Generation FLIR Demonstrator

Parameter	*UFOV*	*NFOV*	*WFOV*	*XFOV*
EFL	24.0″	11.4″	4.36″	2.08″
EPD	4.0″	3.8″	1.45″	0.69″
f/#	6.0	3.0	3.0	3.0
FOV	0.9 × 1.2 deg	0.9 × 2.5 deg	5.0 × 6.6 deg	10.4 × 13.8 deg
Magnification	11.5×	5.5×	2.1×	1.0×
FPA	Third-generation 640 × 480, 20 μm pixel pitch, MWIR & LWIR			

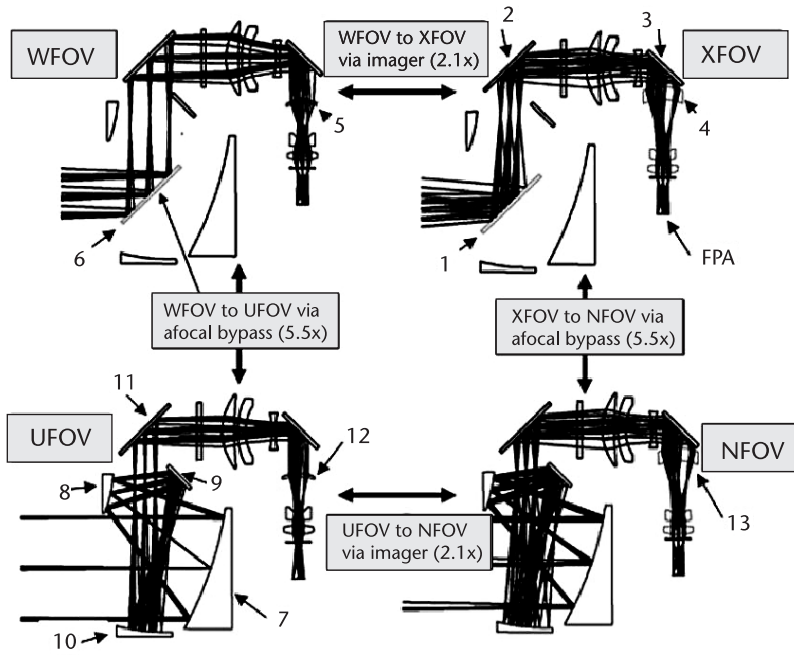


Figure 7.68 Four FOV dual *f*-number optical systems.

the FPA. To go from the UFOV configuration to the NFOV configuration, optical element 12 is mechanically removed and replaced with optical element 13. Optical elements 4 and 13 are identical and optical elements 5 and 12 are identical.

The way the system works can be understood with less detail. The FLIR camera has two FOVs, which are switched by having element 4 or 5 in the optical train. The afocal telescope has three curved mirrors denoted by 7, 8, and 10 with a folding mirror 9. The four FOVs are achieved by combining the two FOVs of the camera with the afocal telescope in and out of the optical system.

The afocal telescope used is an anastigmat. An *anastigmat* is[22] a compound optical system whose astigmatic difference is zero for one or more off-axis zones in the image plane. In such an optical system, the other aberrations are typically sufficiently well corrected to yield excellent definition over the entire field. In an optical system having astigmatism, an off-axis point source is imaged as a line in the sagittal and meridional planes.[10] The *astigmatic difference* is[22] the distance between the sagittal and meridional image planes. An anastigmat is defined more simply in the *American Heritage Dictionary*: A compound lens corrected for astigmatism for at least one off-axis zone in the image plane.

The focal plane array for the system displayed in Figure 7.68 is sensitive to LWIR (8 to 12 μm) and MWIR (3 to 5 μm). If mirror 11 had the property that it reflected LWIR and MWIR but passed visible radiation, then a camera sensitive to visible radiation could be positioned behind mirror 11 and the optical system would be sensitive to visible, MWIR, and LWIR.

There are a great many reflective surfaces and lenses in the optical system of Figure 7.68. To avoid transmission losses, the lens and mirror surfaces need to be treated to avoid reflection losses from the MWIR and LWIR radiation.

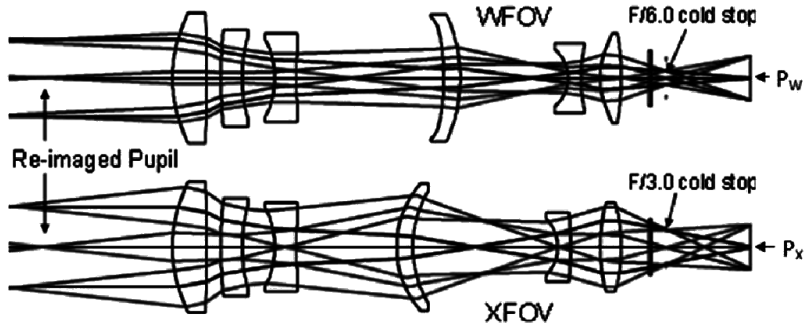


Figure 7.69 Dual *f*-number imager.

Observe from Table 7.1 that the UXOV mode is *f*/6 and the other three modes are *f*/3. As illustrated in Figure 7.69 this is accomplished by changing an aperture stop in the dewar. Arrows indicate the size and location of the *f*/6.0 and *f*/3.0 cold stop. In Figures 7.68 and 7.69, the cold stop is the system aperture stop and also the exit pupil. The image of the aperture stop through the optical elements between the aperture stop and the object is the entrance pupil. In Figure 7.69, the entrance pupil is labeled *Re-imaged pupil*. The multitude of lines in Figure 7.69 is color coded in the original drawing, which facilitates an understanding of the system. The blue rays correspond to a distant point on the optical axis and are brought to a focus at the point labeled P_W and P_X. The green rays correspond to a distant point beneath the optical axis and they are brought to a focus beneath P_W and P_X. Similarly the red rays correspond to a distant point above the optical axis and they are brought to a focus above P_W and P_X. The figures shows three chief rays, each of which goes through the center of the entrance pupil and aperture stop (here labeled cold stop). Note that the pencil of rays converging at P_W has a smaller acceptance angle than those converging at P_X and this is consistent with the *f*-numbers for the two FOVs. Observe that the diameter of the *f*/6.0 cold stop is considerably smaller than the diameter of the *f*/3.0 cold stop. The cold stop is located in the dewar and an ability to change the size of the cold stop is a recent advancement in the state of the art.

7.18 Diffraction Blur

In Section 7.9 we gave expressions for the angle subtended by the blur circle at the center of a lens [see (7.26)] and the size of the blur circle in the image plane [see (7.27)]. These equations assumed a thin lens and a distant object. In this section we extend these results to include nearby objects and thick lenses.

First we describe equations for the diffraction blur size β for a thin lens regardless of where the object is (Figure 7.70). If the object plane is at infinity then the image will be at f and (7.26) applies:

$$\beta = 2.44 \frac{\lambda}{D_{Lens}} \tag{7.79}$$

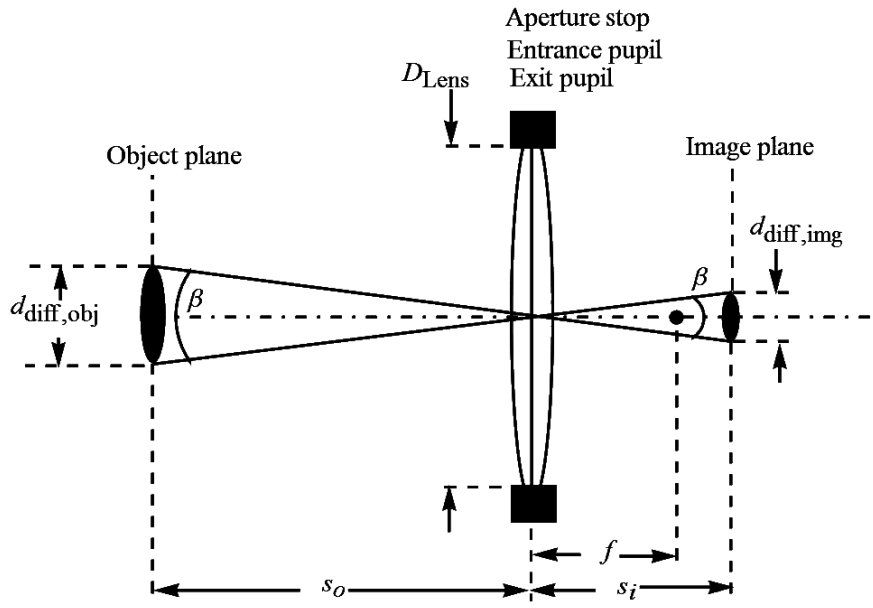


Figure 7.70 Diffraction-limited resolution for thin lens.

If the object plane is at s_o then the image will be at s_i and (7.79) still applies. Since we know β the diameter of the diffraction blur circle in image space is

$$d_{diff,\,img} = \beta s_i \tag{7.80a}$$

where s_i is found from the object position and the Gauss thin lens equation. If the object is at infinity, then $s_i = f$ and (7.80a) becomes

$$d_{diff,\,img} = \beta f \quad \text{when image plane is at infinity} \tag{7.80b}$$

The diameter of the diffraction blur circle in image space is

$$d_{diff,\,obj} = \beta s_o \tag{7.81}$$

We now extend these results to the case of a thick lens. In Figure 7.71 the entrance and exit pupils correspond to those of an arbitrary optical system and can be real or virtual. The angles β_{obj} and β_{img} are the full angles of the diffraction diameter blur circle in object and image space, respectively, as subtended from the center of the entrance and exit pupils. Let D_{enp} and D_{exp} denote the diameters of the entrance and exit pupils with $z_{obj,Enp}$ and $z_{Exp,img}$ being the distance between the object plane and the entrance pupil and the distance between the exit pupil and the image plane, respectively. Then[6]

$$\beta_{obj} = 2.44\frac{\lambda}{D_{Enp}} \tag{7.82}$$

$$\beta_{img} = 2.44\frac{\lambda}{D_{Exp}} \tag{7.83}$$

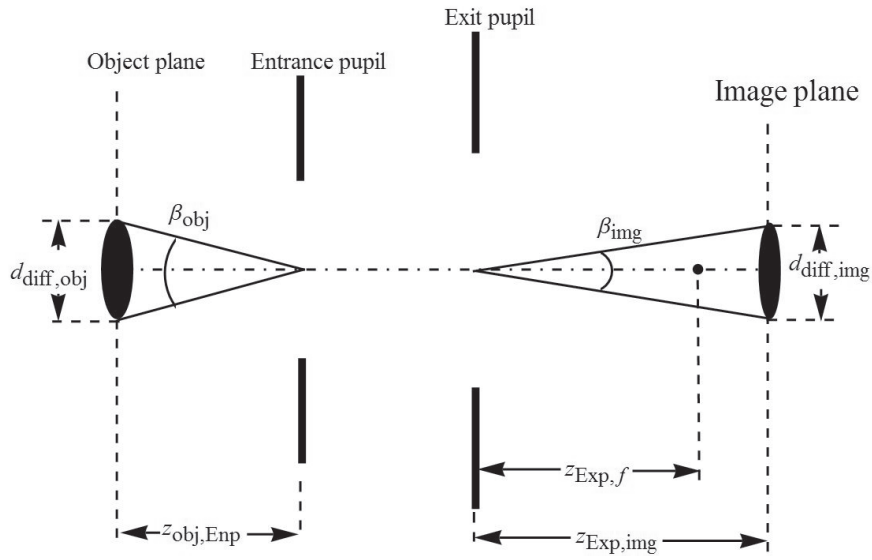


Figure 7.71 Diffraction-limited resolution for thick lens.

The diameters of the entrance and exit pupils will not generally have the same size and so in general $\beta_{obj} \neq \beta_{img}$. The diameters of the diffraction blur spot in the image and object plane are readily calculated[6]:

$$d_{diff,\,img} = \beta_{img} z_{Exp,img} \tag{7.84a}$$

$$d_{diff,\,obj} = \beta_{obj}\, z_{obj,\,Enp} \tag{7.85}$$

If the object is at infinity then the image will be in the focal plane and in that case

$$d_{diff,\,img} = \beta_{img} z_{Exp,\,f} \tag{7.84b}$$

To apply the last three equations, the entrance and exit pupils are found using the techniques described in Section 7.10 and the location of the principal planes is found from the optical design using the ray-tracing rules described in Section 7.3 or a ray-tracing program.

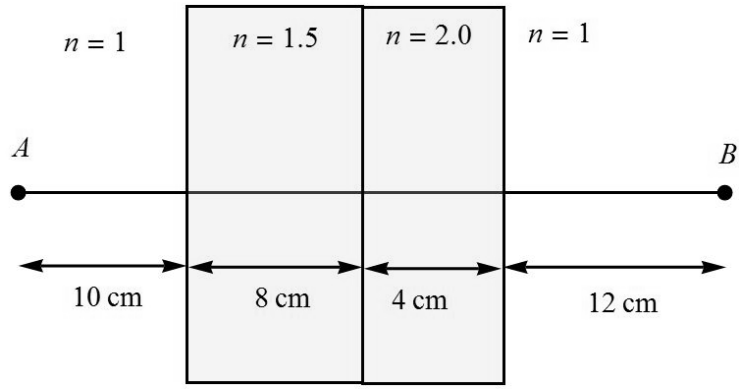
7.19 Guide To the References

Good introductory level optics books with a physics flavor are Refs. 9, 10, 13, 17, 18, and 21. References 1, 3, 12, 19, 23, and 30 deal with optics and have more of an engineering flavor. Consult references 14, 19, 20, 24, and 26 for books dealing with optical engineering. The SPIE field guides 5, 8, and 29 give good compact treatments of their subjects. The five-volume *Handbook of Optics* [2] sponsored by the Optical Society of America provides an encyclopedic treatment of optics. Almost all optics books have some material on aberrations. Reference 15 describes aberration theory in more detail than most of the intro-

ductory texts, whereas Ref. 16 gives an encyclopedic treatment of the subject. Consult Ref. 29 for a compact treatment of binoculars and scopes and Ref. 4 for a fuller treatment of astronomical optics.

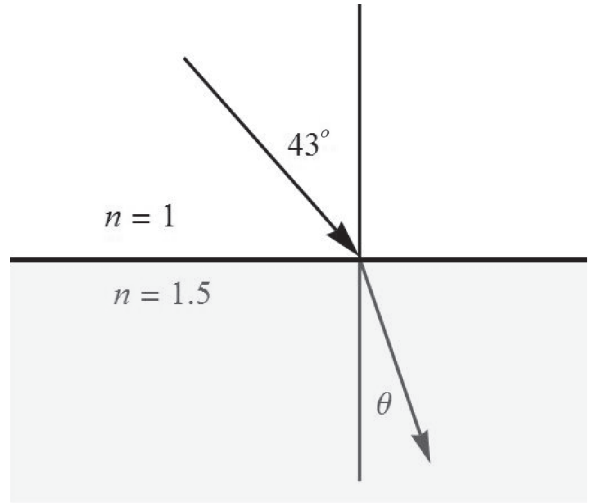
7.20 Exercises

7.1 a) Calculate the OPL from point A to point B in the system shown below.



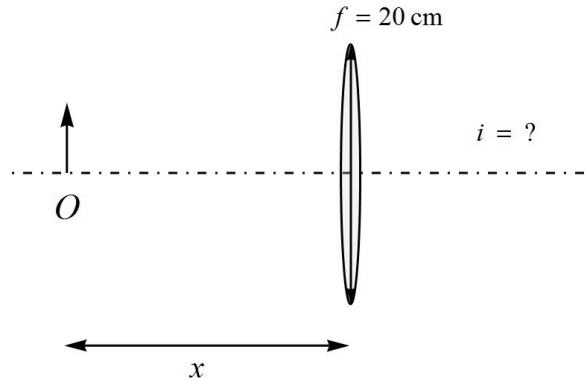
b) Why is the concept of OPL useful?

7.2 Calculate the angle of the refracted ray in the system shown below.

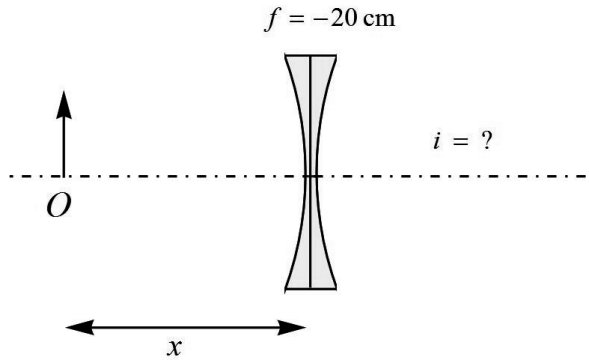


7.3 Determine the image position and size (assume an object size of 1 cm) corresponding to the object shown below when the object distance is
a) $x = \infty$
b) $x = 60$ cm

c) $x = 40$ cm
d) $x = 25$ cm
e) $x = 10$ cm
f) $x = 0.1$ cm



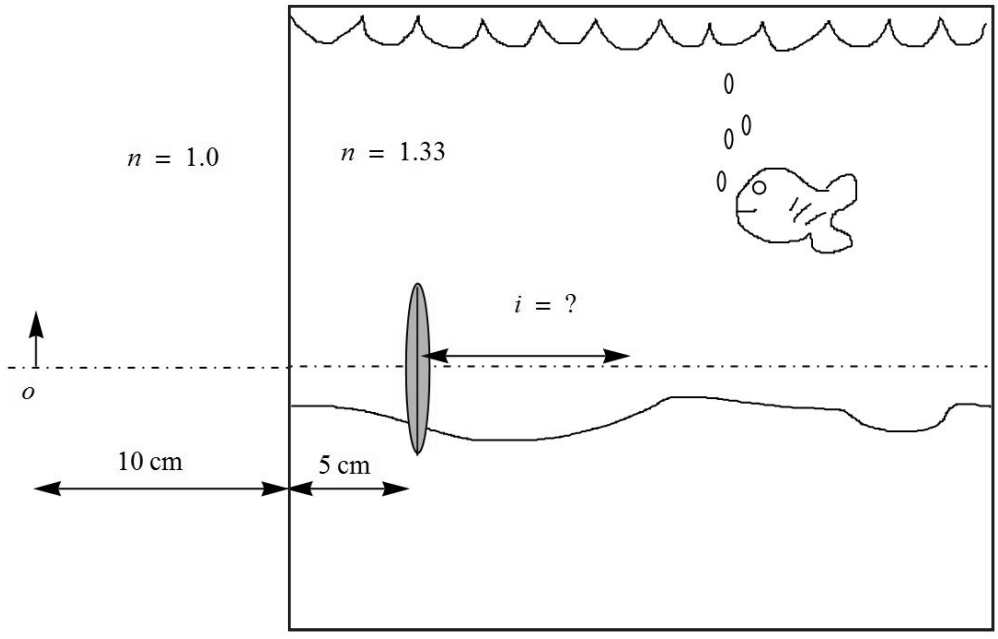
7.4 Determine the image position and size (assume an object size of 1 cm) corresponding to the object shown below when the object distance is
a) $x = \infty$
b) $x = 40$ cm
c) $x = 20$ cm
d) $x = 10$ cm
e) $x = 0.1$ cm



7.5 Calculate the vergence of a wavefront that is 10 cm away from a point source.

7.6 Calculate the vergence of a wavefront that is 30 cm before coming to a focus.

7.7 A lens is dropped upright into a fish tank and is lodged 5 cm from the tank glass in the sand. An object is placed 10 cm from the tank glass. Assuming the glass thickness to be negligible, find the image distance from the lens. The biconvex lens refractive index is 1.67, and the radius of curvature for both sides is 6.8 cm.



7.8 Determine the FOV for a system that has an effective focal length of 80 cm and a detector staring array size of 1.5 cm × 2.0 cm.

7.9 Three lenses with focal lengths of 5, 2, and 5 cm, respectively, are placed 1 cm apart from each other. Given that parallel light enters the first lends, how far behind the third lens does the light come to a focus.

7.10 Determine the diameter of a telescope aperture required to resolve two stars separated by 100 million km where the pair of stars is 10 light-years away from Earth. Assume a wavelength of 0.5 μm.

7.11 Determine the limit on angular resolution for diffraction-limited sensors with apertures of 5, 10, and 15 in. Assume a wavelength of 10 μm.

7.12 Determine both the MTF and psf for a 3- to 5-μm MWIR sensor with a 5-in. aperture and a 10-in. focal length. Specify the MTF in both cycles per milliradian (angular spatial frequency) and cycles per millimeter (spatial frequency in the detector plane). Specify the psf in both angular space and detector plane space.

7.13 A camera lens with a diameter of 50 mm has a focal length of 100 mm. The field stop in the horizontal direction is 35 mm. Assume the camera lens can be approximated by a thin convex lens.
a) What is the *f*-number of the lens?
b) What is the horizontal FOV when the camera is focused at infinity? At 200 mm?

c) What is the working *f*-number of a lens when it is focused at infinity and when it is focused at 200 mm and how does that compare with the *f*-number given by (7.54)?

7.14 A telescope has an objective lens with a focal length of 20 cm, an eyepiece lens with a focal length of 5 cm, and is interfaced to a camera with a 50-mm focal length lens. The diameter of the objective is 14 cm. Assume the camera is intended for use with visible light.
a) What is the angular magnification of the telescope?
b) What is the maximum angular magnification of the telescope before empty magnification sets in and what eyepiece focal length does this correspond to?
c) What is the diffraction angular resolution of the telescope?
d) What is the diameter of the exit pupil of the telescope and where is it located?
e) What is the *f*-number of the telescope-camera system?

7.15 Suppose the imager illustrated in Figure 7.69 is to be modified for close-in airport security. The engineer is considering axially moving the cold stop so the entrance pupil moves closer to the subject (i.e., $z_{obj,\ Enp}$ is decreased). As the entrance pupil moves closer to the subject, it appears from Figure 7.69 that the entrance pupil diameter will remain constant, which implies [see (7.82)] that β_{obj} will remain constant. Then (7.85) implies the diffraction blur spot on the subject will be smaller, and this implies the diffraction limit has been improved without increasing the diameter of the objective lens in Figure 7.69. Discuss the validity of this argument.

7.16 Use the concept of optical path length to show that it is possible to design an optical system that according to the laws of geometric optics images a point source in air to a fixed location in a point in glass. What is the equation of the air-glass surface that has this property? A surface that does this is called a Cartesian oval. [10]

7.17 Look up the definition of a zoom lens in the *Photonics Dictionary*. It says a zoom lens is an optical system of variable focal length with the focal plane remaining in a fixed position. As the focal length of a lens is increased, the field of view decreases [see (7.19)]. Thus a zoom lens trades FOV to see better detail in a smaller part of the scene and it does this by putting more detector elements on the target. However, sometimes the difficulty in performing a task is not having a high enough SNR. Design an electro-optical system that trades FOV view for improved SNR. Reference US Patent No. 6774366, http://www.freepatentsonline.com/6774366.pdf.

References

[1] Banerjee, P. P., and T. C. Poon, *Principles of Applied Optics*, Boston, MA: Aksen Associated, (IRWIN), 1991.

[2] Bass, M., *Handbook of Optics, Vol. 1: Geometrical and Physical Optics, Polarized Light, Components and Instruments*, New York City, NY: McGraw-Hill, 2010.

[3] Boreman, G. D., *Basic Electro-Optics for Electrical Engineers*, Bellingham, WA: SPIE Press, 1998.

[4] Breckinridge, J. B., *Basic Optics for the Astronomical Sciences*, Bellingham, WA: SPIE Press, 2011.

[5] Daniels, A., *Field Guide to Infrared Systems, Detectors and FPAs*, Bellingham, WA: SPIE Press, 2010.

[6] Dereniak, E. L., and G. D. Boreman, *Infrared Detectors and Systems*, Hoboken, NJ: Wiley, 1996.

[7] Gaskill, J. D., *Linear Systems, Fourier Transforms, and Optics*, Hoboken, NJ: Wiley, 1978.

[8] Greivenkamp, J. E., *Field Guide to Geometrical Optics*, Bellingham, WA: SPIE Press, 2004.

[9] Guenther, R., *Modern Optics*, Hoboken, NJ: Wiley, 1990.

[10] Hecht, E., and A. Zajac, *Optics*, Boston, MA: Addison-Wesley, 2001.

[11] Holst, G. C., *Electro-Optical Imaging System Performance*, Bellingham, WA: SPIE Press, 2008.

[12] Iizuka, K., *Engineering Optics*, New York City, NY: Springer-Verlag, 2008.

[13] Jenkins, F. A., and H. E. White, *Fundamentals of Optics*, New York City, NY: McGraw-Hill, 2001.

[14] Kingslake, R., and R. B. Johnson, *Lens Design Fundamentals*, Bellingham, WA: SPIE Press, 2010.

[15] Mahajan, V. N., *Aberration Theory Made Simple*, Bellingham, WA: SPIE Press, 2011.

[16] Mahajan, V. N., *Imaging and Aberrations Parts I and II*, SPIE Press, 1998, 2001.

[17] Meyer-Arendt, J. R., *Introduction to Classical and Modern Optics*, Upper Saddle River, NJ: Prentice-Hall, 1984.

[18] Moller, K. D., *Optics*, Mill Valley, CA: University Science Books, 1988.

[19] Mouroulis, P., and J. Macdonald, *Geometrical Optics and Optical Design*, Oxford University Press, 1997.

[20] O'Shea, D. C., *Elements of Modern Optical Design*, Hoboken, NJ: Wiley, 1985.

[21] Pedrotti, F. L., and L. S. Pedrotti, *Introduction to Optics*, Prentice Hall, 1993.

[22] Photonics Dictionary, Laurin Publishing Co, (1996) http://www.photonics.com/Directory/Dictionary/Default.aspx

[23] Pinson, L, J., *Electro-Optics*, Hoboken, NJ: Wiley, 1985.

[24] Riedl, M. J., *Optical Design: Applying the Fundamentals*, Bellingham, WA: SPIE Press, 2009.

[25] Saleh, B. H., and M. C. Teich, *Fundamentals of Photonics*, Wiley, 2007.

[26] Smith, W. J., *Modern Optical Engineering*, New York City, NY: McGraw-Hill, 2000,

[27] Vizgaitis, J., "Third generation infrared optics," *Proc. SPIE,* Vol. 6940, 2008.

[28] Vizgaitis, J., J. Miller, J. Hall, and D. Berube, "3rd Generation FLIR Demonstrator," *Proc. SPIE, Vol.* 6940, 2008.

[29] Yoder, P. R., and D. Vukobratovich, *Field Guide to Binoculars and Scopes*, Bellingham, WA: SPIE Press, 2011.

[30] Yu, F. T., and X. Yang, *Introduction to Optical Engineering*, New York City, NY: Cambridge University Press, 1997.

[31] Vizgaitis, J., Briefing notes.

[32] McCluney, R., *Introduction to Radiometry and Photometry*, Norwood, MA: Artech House, 1994.

CHAPTER 8

Detectors

The detector is the component of an optical system that transforms the optical signal into an electrical signal. This electrical output is proportional to the incident optical power. The detector component in the sensor plays a key role in determining system-level parameters including spectral operating band, sensitivity, and resolution. The spectral response is determined by the detector material characteristics and the operating temperature. A strong energy band for photon acceptance must be present for the energy conversion process to occur. The detector sensitivity is a function of material (i.e., energy gap), wavelength, detector size, bandwidth, and shielding. Finally, the resolution is set by the detector size and pitch (center-to-center spacing) and the optical focal length.

The detectors must be distributed over the image plane in order to convert the image into a spatial signal. The spatial representation can be performed in terms of a continuous signal or in terms of a sampled signal. The detector distribution across the entire image is accomplished by scanning the image across single detectors or linear detector arrays. The spatial distribution is inherent with staring arrays. We focus on the spectral sensitivity and detector resolution in this chapter. Since the first edition of this book, staring arrays have been used widely and the majority of new systems include staring arrays. There are many new detector materials and arrays in the infrared and shortwave infrared regions and we have added sections to this chapter to describe the new detectors.

8.1 Types of Detectors

There are two general classes [1, 2] of detectors: *photon* (or quantum) and *thermal* detectors. Photon detectors convert absorbed photon energy into released electrons (from their bound states to conduction states). The material bandgap describes the energy necessary to transition a charge carrier from the valence band to the conduction band. The change in charge carrier state changes the electrical properties of the material. These electrical property variations are measured to determine the amount of incident optical power. Thermal detectors absorb energy over a broad band of wavelengths. The energy absorbed by a detector causes the temperature of the material to increase. Thermal detector materials have at least one inherent electrical property that changes with temperature. This temperature-related property is measured electrically to determine the power on the detector. Figure 8.2 shows the detectors described in this chapter.

Photon detectors include photoconductors, photovoltaic (PV), and photoemissive detectors. The photoconductive detector is constructed of semiconductor material that converts absorbed photons into free-charge carriers. The change in free carriers varies the conductivity (and resistance) of the detector. A simple circuit provides a bias current that reflects the change in conductivity.

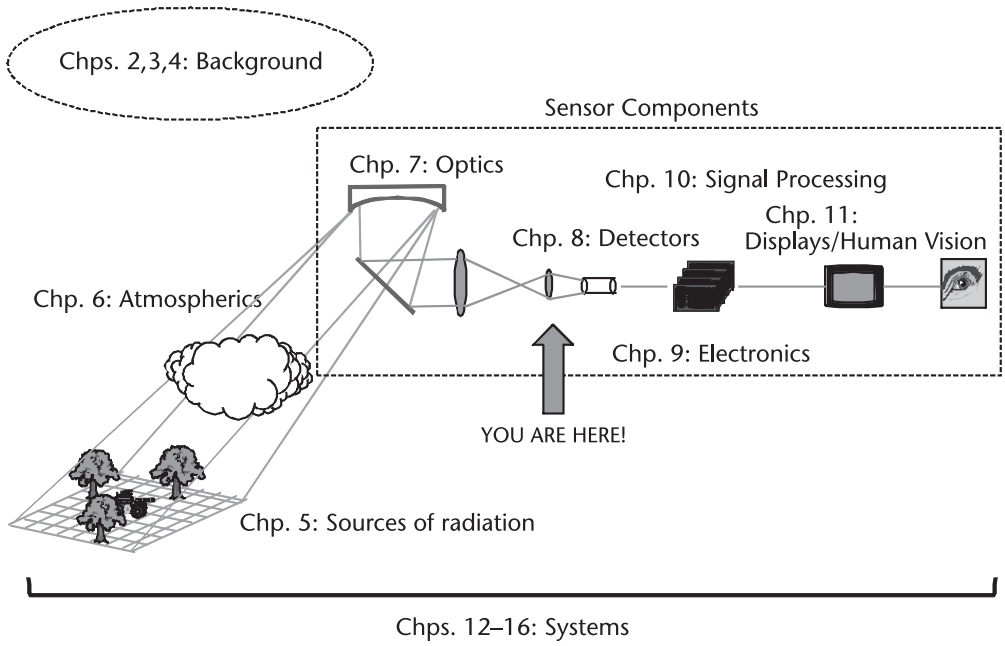


Figure 8.1 Detector and scanner components of the system.

The PV detector absorbs photons at a P-N junction. Hole-electron pairs are generated that alter the junction voltage. The change in voltage describes the amount of optical power on the detector. The photovoltaic detector does not require any external bias voltage. The inherent voltage of the P-N junction changes without bias and can be read out directly. Like photoconductors, the PV detector has spectral characteristics that are derived from the material's molecular energy levels.

Photoemissive devices are different from photoconductors and PV detectors in that an electron physically leaves the detection material when a photon is absorbed. Usually placed in a vacuum tube with high voltage, a photocathode absorbs the photon and the energy absorbed is enough (higher than the electron work function) to release the electron in the vacuum chamber. The voltage potential moves it toward the anode. Once electrons reach the anode, they are measured as a current that reflects the optical flux on the photocathode.

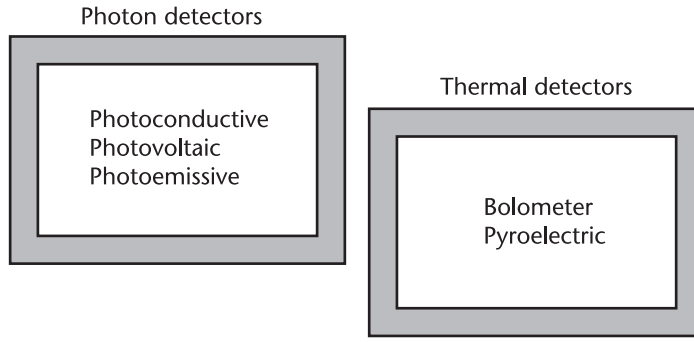


Figure 8.2 Types of detectors.

Thermal detectors [3] include bolometers and pyroelectric detectors. Bolometers are sometimes called thermistors, because the material provides a resistance that changes with temperature. Semiconductors are used with high-temperature coefficients of resistance. The material is usually a long, narrow strip of material coated with highly absorbent coating. The electrical resistance of the strip is measured with a bias current to determine the amount of absorbed radiation.

Pyroelectric detectors used to be the more common thermal detectors, but the past decade has seen wide proliferation of microbolometer arrays. The pyroelectric detector is a ferroelectric material between two electrodes. The material exhibits a change in polarization with temperature. It can be thought of as a capacitor whose capacitance is temperature sensitive. As with any capacitor, measurement requires an alternating current signal, so most pyroelectric detectors are chopped to vary the incident flux.

In this chapter, we discuss each of the photon and thermal detectors mentioned above. The list is not exhaustive, but it does contain most of the commonly specified detectors.

8.2 Photon Detectors

Photon detectors generate free-charge carriers through the absorption of photons. The absorption occurs without any significant increase in the detector temperature [4]. Detection occurs by a direct interaction of the photon with the atomic lattice of the material. Material parameters that can be changed by the interaction are resistance, inductance, voltage, and current. The concepts presented here apply to detectors that sense wavelengths from 0.2 to 1,000 μm.

A variety of bandgaps (and corresponding wavelength sensitivities) can be achieved with the selection of materials. Recall that $E = h\eta$, where $\nu = c/\lambda$, h is Planck's constant, and c is the speed of light. The wavelength of a photon has to be short enough so that the photon energy exceeds a material's bandgap energy level for absorption. Intrinsic detectors are those that are constructed from semiconductor crystals with no introduced impurities or lattice defects [5]. Extrinsic detectors have impurities introduced to change the effective bandgap energy.

Cryogenic cooling (with closed-cycle coolers or liquid nitrogen) is typically required for infrared photon detectors to reduce dark-current generated electron noise. The reduction in temperature reduces the level of dark-current noise to below that of a photon-generated signal. Photon detector responsivity is sometimes characterized by a quantum efficiency that describes the number of electron transitions per absorbed photon. Intrinsic detectors have quantum efficiencies of around 60%. Typical quantum efficiencies for photoconductors range from 30% to 60%. Photovoltaic and photoemissive detectors have quantum efficiencies of around 60% and 10%, respectively.

The detection mechanisms for the three most common types of photon detectors are covered in this chapter. The following sections describe photoconductive, photovoltaic, and photoemissive detectors.

Photoconductors

Photoconductors sense optical power with a change in conductance (or inverse resistance). Longer wavelengths can be detected by *doping* a pure material with impurities, giving an extrinsic detector. The energy gap transition for an intrinsic detector is depicted in Figure 8.3(a). An extrinsic bandgap is shown in Figure 8.3(b). A P-type extrinsic detector decreases the conduction band energy level, and an N-type extrinsic detector raises the valence band energy level. The result is a smaller bandgap material that can detect longer wavelengths. Quantum efficiencies are lower (around 30%) for extrinsic detectors because the dopant is less abundant than the host material.

For photoconductors, an increase in the incident optical power increases the number of free-charge carriers, so the effective resistance of the detector decreases. The differential charge provided by the impinging photons is given by [1]

$$dq = E_p \eta_q w t_q e I_d \quad [\text{C}] \tag{8.1}$$

where E_p is the incident irradiance in photons per second per square centimeter, w is the detector width in centimeters, l_d is the detector length in centimeters, η_q is the quantum efficiency, t_q is the charge carrier mean lifetime in seconds, and e is the charge on an electron (1.602×10^{-19} C). Figure 8.4 shows a typical photoconductor.

The photo-induced current is related to the transit time of the charge carrier across the detector length l_d, where the transit time is estimated to be $dt = (l_d)^2/(\mu V_{bias})$, where V_{bias} is the bias voltage across the detector and μ is the carrier mobility in cm^2/(V-s). Now the photo-induced current is

$$i_s = \frac{E_p \eta_q w t_q e \mu V_{bias}}{l_d} \quad [\text{A}] \tag{8.2}$$

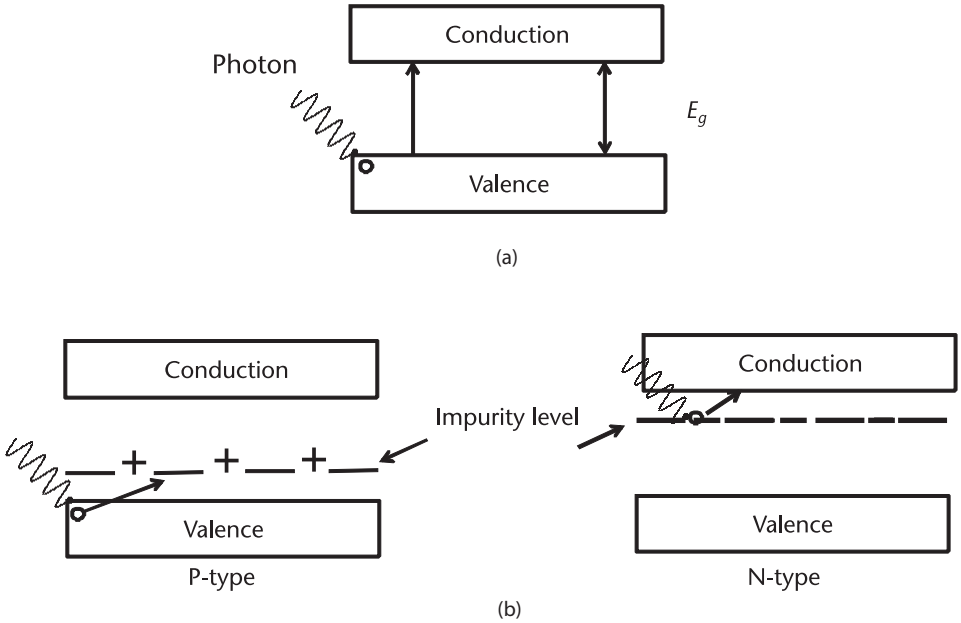


Figure 8.3 (a) Intrinsic and (b) extrinsic detection.

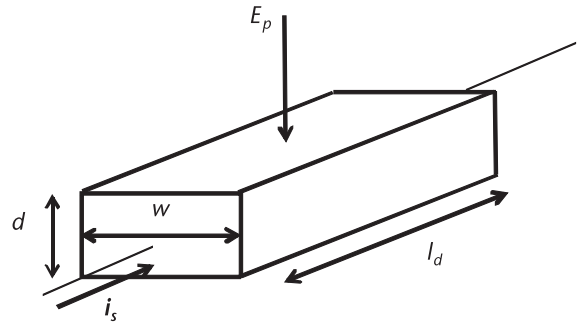


Figure 8.4 Photoconductive detector.

The photo-induced change in resistance is the bias voltage divided by the photo-induced current:

$$\Delta R_s = \frac{l_d}{E_p \eta_q w t_q e \mu} \ [\Omega] \tag{8.3}$$

Finally, a photo-induced conductivity can be found using the conductivity-resistance relationship:

$$\Delta R_s = \frac{l_d}{\Delta \sigma_s w d} \ [\Omega] \tag{8.4}$$

where $\Delta\sigma_s$ is the photo-induced conductivity of the resistor.

Example 8.1

An indium antimonide photoconductor detector is 100 μm long and 50 μm wide. Its quantum efficiency is 0.45 and its mean carrier lifetime is 5×10^{-7} seconds. Its mobility is 3×10^4 cm^2/(V-s). Determine the detector's change in resistance for an in-band incident irradiance of 3×10^{17} photon/(s-cm^2). The solution is given by (8.3):

$$\Delta R_s = \frac{(0.01)}{\left(3\times10^{17}\right)(0.45)(0.005)\left(5\times10^{-7}\right)\left(1.602\times10^{-19}\right)\left(3\times10^{4}\right)} = 6.2\text{k}\Omega$$

The conductivity is found, using (8.4), to be 0.064 S/cm.

Photovoltaic Detectors

Photovoltaic detectors are constructed with the formation of a P-N type junction in a semiconductor. The P-N junction has an inherently generated voltage and does not require a bias voltage or current. A typical photovoltaic detector energy diagram is shown in Figure 8.5.

The photons are absorbed at the junction of the P- and N-type materials, an area called the *depletion region*. The junction is the same as a diode with the added effect that voltage is generated with illumination. The energy diagrams of Figure 8.5 show the diode in the dark and illuminated states. The voltage in the dark state is related to the differences in conduction (and corresponding valence) energy levels

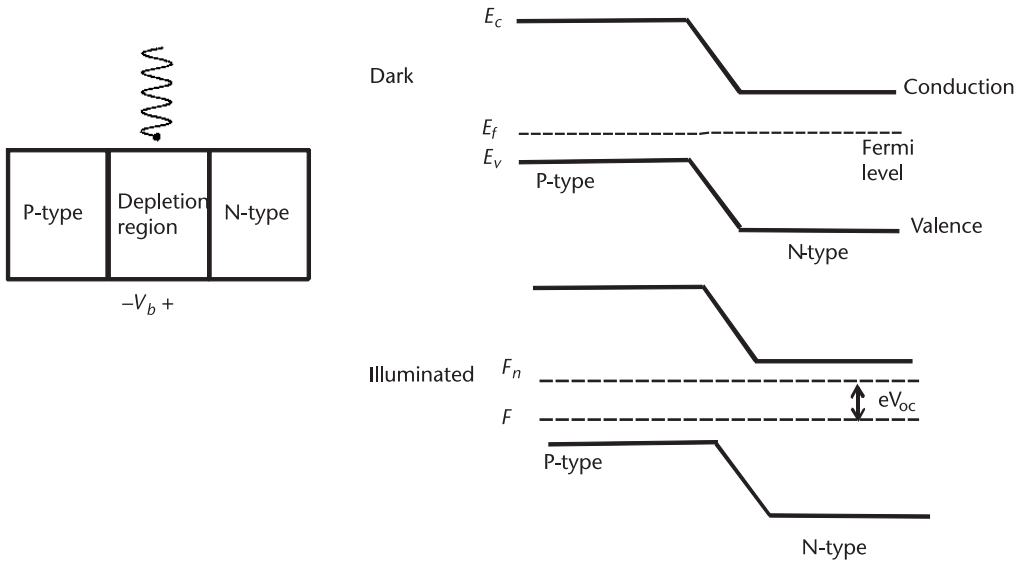


Figure 8.5 Photovoltaic energy diagram.

for the P- and N-type materials across the depletion region. Pinson [1] provides the following voltage description. The voltage is given by

$$V_b = \frac{kT}{e} ln \frac{n_n}{n_p} \quad [\text{V}] \tag{8.5}$$

where k is Boltzmann's constant (1.38×10^{-23} J/K or 8.62×10^{-5} eV/K), T is the junction temperature in kelvin, e is the charge on an electron, and n_n and n_p are the free-electron densities in the N- and P-type materials, respectively. The Fermi level shown describes the energy at which 50% of the energy levels are likely to be occupied by electrons. This energy level applies when no excess carriers are generated by optical excitation. Quasi-Fermi levels apply for the optically excited state, where these levels describe how far the electron and hole populations have deviated from the equilibrium values. The open-circuit voltage is proportional to the difference in these energy levels.

The voltage across the junction when excess electron-hole pairs are generated can be described by the changes in the electron and hole populations, δn_n and δn_p. The junction voltage becomes

$$V_b = \frac{kT}{e} ln \frac{n_n + \delta n_n}{n_p + \delta n_p} \quad [\text{V}] \tag{8.6}$$

The changes in the number of excess electron-hole pairs can be approximated by

$$\delta n_n = \delta n_p = \frac{n_q E_s}{d} t_n \quad [\text{electron/m}^3] \tag{8.7}$$

where d is the detector depth and t_n is the carrier lifetime. A differential incident photon irradiance gives a differential voltage of

$$\frac{dV_b}{dE_s} = \frac{d}{dE_s}\left[\frac{kT}{e} ln \frac{n_n + \frac{n_q E_s}{d} t_n}{n_p - \frac{n_q E_s}{d} t_n}\right] \approx \frac{kT}{e}\left[\frac{\frac{n_q}{d} t_n \left(n_n + n_p\right)}{\left(n_p + \frac{n_q E_s}{d} t_n\right)\left(n_p - \frac{n_q E_s}{d} t_n\right)}\right] \tag{8.8}$$

If the number of excess electron-hole pairs generated by the irradiance can be assumed to be small compared with n_n, then

$$\frac{dV_b}{dE_s} \approx \frac{kTn_q t_n (n_n + n_p)}{edn_n n_p} \tag{8.9}$$

Pinson states that this differential quantity is a constant that relates the photo-induced voltage to the incident irradiance. The current-voltage characteristics of photovoltaic detectors are very similar to those of a diode. Because of this similarity, photovoltaic detectors are sometimes referred to as photodiodes. Photovoltaic detectors offer several advantages over photoconductive detectors including better responsivity, simpler biasing, and a better theoretical SNR.

The photovoltaic detector is normally operated in the open-circuit mode, where the photo-induced voltage is measured directly. This mode has SNR advantages over a system that requires external amplification. However, the detector can be operated in the reverse-bias direction, known as the *photoconductive mode*. The electric field strength and the depletion region width are increased. Advantages are higher speed, lower capacitance, and better linearity. The disadvantage is that dark currents are greater.

The avalanche photodiode (APD) [6] is a specialized photovoltaic detector that is operated in the photoconductive mode. Large reverse-bias voltages are applied to the P-N junction. This causes some of the electrons passing through the large electric field to gain added energy so that additional electron-hole pairs are generated. This process is known as *impact ionization*. If the newly created electron-hole pairs attain sufficiently high energy levels, they create more electron-hole pairs. This process is known as *avalanche multiplication* and is the means for large internal gains. Typical internal gains are from 50 to 500.

Photoemissive Detectors

Photoemissive detectors release electrons from the photosensitive surface during the detection process. The incident photon of energy $h\nu$ provides the energy necessary to release an electron from the surface. The process is shown in Figure 8.6.

The photon enters a glass envelope that houses the photocathode and anode. Usually the envelope comprises a vacuum tube where conductive material deposited on the glass surface acts as the cathode. The photon energy is absorbed, producing a free electron that is accelerated to the anode through the vacuum. Only electrons that have sufficient energy to overcome the work function of the photocathode material result in released electrons. The electrons are collected at the anode and the corresponding current is measured to determine the amount of light on the photocathode. Photocathodes are designed with high optical absorption and small work functions.

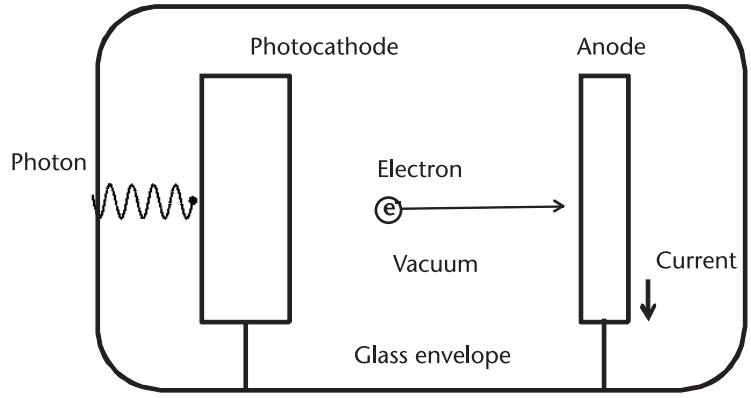


Figure 8.6 Photoemissive detection.

The photomultiplier tube (PMT) is a special case of the photoemissive detector. The PMT is used when large gains are required to sense very low light levels. The PMT consists of a photocathode, electron multipliers (dynodes), a collection anode, and an evacuated glass or metal envelope. The photomultiplier is shown in Figure 8.7.

An electron is released from the photocathode and is accelerated to the first dynode. The electron strikes the first dynode with enough energy to release a number of electrons that are accelerated to the next dynode. This multiplication occurs through a number of dynodes resulting in a large number of electrons collected on the anode. The gain of a PMT is estimated by

$$G = \chi^n \tag{8.10}$$

where G is the current gain, χ is the secondary emission ratio for dynodes, and n is the number of dynodes. A nine-stage PMT with a secondary emission ratio of 8 gives a gain of around 10^7. This large gain and low noise make the PMT an attractive detector for many applications.

8.3 Thermal Detectors

For thermal detectors, the detector signal is related to the temperature of the detector material, which is altered because of incident radiation. These detectors typically

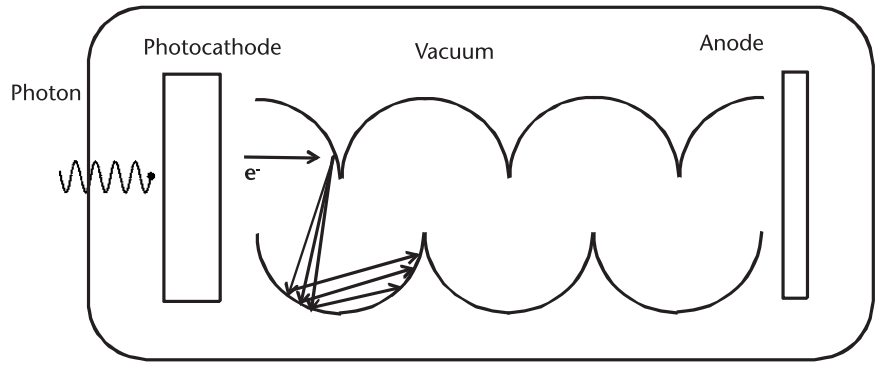


Figure 8.7 Photomultiplier.

operate at room (ambient) temperature and do not require cryogenic cooling. The change in temperature produces one of three electrical property changes in the detector material: resistance, voltage, or capacitance. Thermal detectors are grouped by the type of electrical property change.

In the past, thermal detectors have been characterized by their slow response time. Significant developments in recent years, however, have made these detectors viable candidates for many applications. Thermal detectors today exhibit many excellent characteristics, including quick response time, low noise, and linear operation. Although the sensitivity of thermal detectors has improved dramatically in recent years, they are still behind the performance of photon detectors. In this section, we discuss two common types of thermal detectors for imaging sensors: the bolometer and pyroelectric detector.

Bolometers

Bolometer detection is derived from a change in the resistance of the detector material. A simple bolometer detector consists of a thin, blackened metal or semiconductor filament, as shown in Figure 8.8.

The change in the bolometer resistance due to incident radiation is given by [7]

$$\Delta R_b = R_o \alpha \Delta T \quad [\Omega] \tag{8.11}$$

where ΔT is the change in temperature of the filament, R_o is the strip resistance when $\Delta T = 0$, and α is the temperature coefficient of resistance. The temperature coefficient is positive for metals and negative for semiconductors. In the absence of incident radiation, there are changes in the detector from conduction from the surrounding environment. The changes are statistical in nature and contribute to detector noise. For this reason, the detector is isolated from thelocal environment to help ensure that signals are a result only of incident radiation. Isolation is achieved

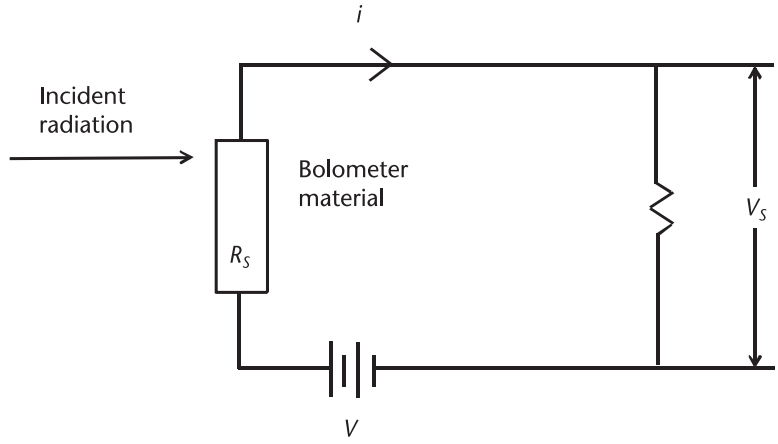


Figure 8.8 Bolometer configuration.

through special manufacturing techniques and by placing the bolometric detector in a vacuum.

The incremental temperature difference is defined by

$$\Delta T = \frac{\Delta\Phi}{\left[K^2 + \left(2\pi f_c C_T\right)^2\right]^{1/2}} \quad [\mathrm{K}] \tag{8.12}$$

where $\Delta\Phi$ is the absorbed radiant power in watts, K is the heat transfer coefficient in Watts per Kelvin, f_c is the chopping signal frequency in Hertz (if the detector is chopped), and C_T is the heat capacity of strip in joules per Kelvin.

The voltage responsivity is the ratio of the change in voltage with respect to the change in input radiation:

$$R = \frac{\Delta V_s}{\varepsilon\Delta\Phi} = \frac{iR_o\alpha\,\varepsilon}{\left[K^2 + \left(2\pi f_c C_T\right)^2\right]^{1/2}} \quad [\mathrm{V/W}] \tag{8.13}$$

where ε is the detector material emissivity. At low frequencies, $f_c << 1/2\pi\,(C_T/K)$, (8.13) can be simplified to

$$R = \frac{iR_o\alpha\,\varepsilon}{K} \quad [\mathrm{V/W}] \tag{8.14}$$

Example 8.2
Calculate the responsivity for a metal bolometer at low frequencies given the following parameters:

i = 15 mA
$K = 5 \times 10^{-4}$ (W/K)
R_o = 50 Ω
α = 0.01/K
ε = 0.9

Inserting the values into (8.14), R is determined to be:

$$R = \frac{\left(15\times10^{-3}A\right)\left(\frac{50V}{A}\right)\left(\frac{0.01}{K}\right)(0.9)}{\left(5\times10^{-4}W/K\right)} = 1.35 \ [\mathrm{V/W}] \tag{8.15}$$

Pyroelectric Detectors

Pyroelectric detectors have a change in surface charge with a change in temperature. This pyroelectric effect, as it is called, produces a current that is proportional to the change in temperature:

$$i_p = p_T\ A_d \frac{dT}{dt} \quad [\mathrm{A}] \tag{8.16}$$

where p_T (C-cm^{-2}-K^{-1}) is the pyroelectric coefficient at the temperature T Kelvin, A_d is the surface area of the detector in square centimeters, and dT/dt is the time rate of temperature change. When a constant radiation is applied to the detector, no current can be sensed. A chopper must be used with the detector in order to provide changing temperatures that produce a current to sense a constant signal. Pyroelectric detectors are in essence capacitors. The responsivity for a pyroelectric detector is similar to that of the bolometer and is given by

$$R = \frac{\Delta V_s}{\varepsilon \Delta \Phi} = \frac{4\pi \varepsilon \rho A_d f_c}{\left[K^2 + \left(2\pi f_c C_T\right)^2\right]^{1/2}} \quad \text{[V/W]} \tag{8.17}$$

where ρ is the pyroelectric coefficient, A_d is the detector area, and f_c is the chopping frequency.

Pyroelectric materials are ferroelectric crystals that show spontaneous changes in polarization along one axis. Two example materials used for this kind of detector are PZT (doped lead zirconate titanate) and BST (barium strontium titanate) [8]. The polarization of these materials varies with the temperature and is characterized by the pyroelectric coefficient. This family of thermal detectors is characterized by fast response times across a broad spectrum. With the success of microbolometers in the past 10 years, it is becoming less common to see pyroelectric detectors in infrared systems.

8.4 Charge-Coupled Devices

The charged-coupled device (CCD) [9–11] has become a very familiar term due to the popularity of handheld video cameras operating in the visible band, although there are also infrared CCDs (IRCCDs). The term *charged-coupled device* does not refer to a specific detector type but rather a detector readout architecture. The basic concept of charge coupling is that once the photo-generated charge is collected in potential wells, they are selectively transferred or read out of each detector location in a serial process. CCDs are typically metal-oxide structure (MOS) devices. The basic structure for these gates is shown in Figure 8.9.

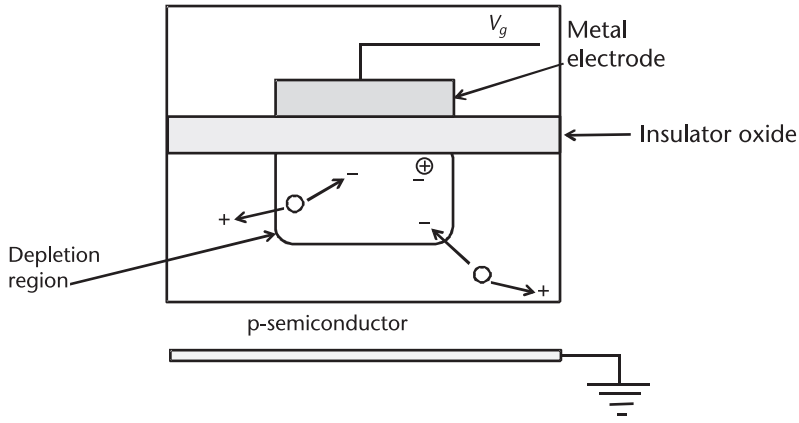


Figure 8.9 Metal-oxide structure.

When a positive voltage is applied, V_g, between the top metal electrode and the substrate, the majority carriers (holes) are moved from the oxide layer creating a depletion region at the interface. If a photon is absorbed into the region, then an electron-hole pair is produced. For the photon to be absorbed, its energy must be greater than that of the bandgap.

The CCD is an analog shift register in which an entire sequence of charge packets can be shifted simultaneously along the surface of the chip by applying clock voltages (V_g). The electric charge varies from one packet to the next. The sequence of charge packets is scanned as the packets move past an output electrode connected to an on-chip amplifier. To read out the charge packets, each component in the array is connected to sequential shift registers on the periphery. The packets are then clocked out of the shift registers by a clock that controls the time delay.

As can be expected, the readout process is not perfect and, hence, a small amount of charge is lost with each transfer. This loss is quantified by the charge-transfer inefficiency η_{ct}. For a total of P charge transfers, the total fractional loss is given by $a = 1 - (1 - \eta_{ct})^P \cong P\eta_{ct}$. The typical charge inefficiency is on the order of 10^{-4}, and typical array sizes (i.e., charge transfers) are $1{,}000 \times 1{,}000$ elements.

8.5 Detector Responsivity

The fundamental purpose of a detector is to convert an optical signal into an electrical current or voltage. The term *responsivity* is used to describe the amplitude of the electrical signal with respect to the incident flux. Responsivity is a function of wavelength and temporal frequency and can be determined by

$$R(\lambda, f) = \frac{V_{out}}{\Phi_{incident}} = \frac{V_{out}}{E(\lambda, f)A_d} \quad \text{[V/W or V/L]} \tag{8.18}$$

where V_{out} is the detector output voltage, $\Phi_{incident}$ is the incident flux in watts, $E(\lambda, f)$ is the incident irradiance in watts per square centimeter, and A_d is the detector area in square centimeters. Current responsivity would have units of Amperes per Watt instead of volts per watt. A typical photon detector responsivity curve is shown in Figure 8.10.

The responsivity is typically measured by the detector manufacturer and provided as a specification or calibration document. The band-averaged responsivity over a spectral band $[\lambda_1, \lambda_2]$ is given by

$$R = \frac{\int_{\lambda_1}^{\lambda_2} \Phi(\lambda) R(\lambda)\, d\lambda}{\int_{\lambda_1}^{\lambda_2} \Phi(\lambda)\, d\lambda} \quad \text{[V/W or A/W]} \tag{8.19}$$

This value is used frequently in cases where the incident flux and the responsivity are not expected to vary significantly over the spectral band. For EO systems, the responsivity is given [10] frequently in terms of photometric units:

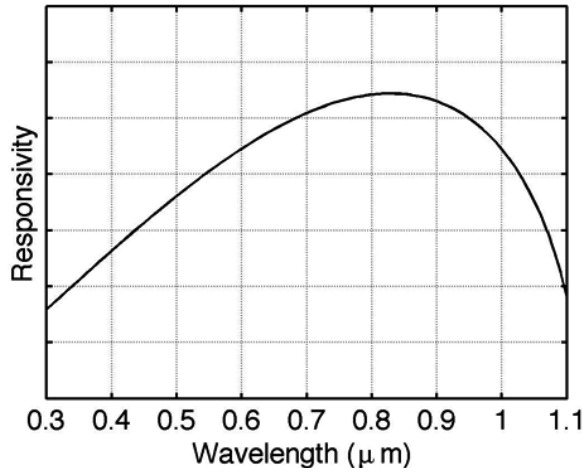


Figure 8.10 Responsivity curve.

$$R_{photopic} = \frac{\int_{\lambda_1}^{\lambda_2} \Phi(\lambda) R(\lambda)\, d\lambda}{683 \int_{\lambda_1}^{\lambda_2} V(\lambda)\Phi(\lambda)\, d\lambda} \quad \text{[V/lumen or A/lumen]} \tag{8.20}$$

where $V(\lambda)$ is the human eye's photopic efficiency. There are 683 L/W at the peak visual efficiency. For EO systems that are closely matched to the human eye in spectral response, the use of photometric responsivity is reasonable. For systems with other-than-photopic spectral responses, however, large differences can be seen between actual voltages and those predicted by the photopic responsivity.

Example 8.3

The irradiance on a detector at a wavelength of 10.6 μm is 100×10^{-6} W/cm^2. The detector has a surface area of 5 cm^2 and the output voltage is 35 mV. Determine the detector's responsivity. The solution is determined by calculating the power on the detector. It is equal to $(100 \times 10^{-6})(5) = 5 \times 10^{-4}$ W. The responsivity is then $R = 0.35/5 \times 10^{-4} = 70$ V/W.

In a manner similar to that of responsivity, the flux response of infrared systems is typically characterized by the system intensity transfer function (SITF). The SITF is measured with a large square target against a uniform background as shown in Figure 8.11(a).

The temperature difference between the target and the background (ΔT) is varied, and the corresponding output voltage is recorded. A typical SITF curve is shown in Figure 8.11(b). The SITF value is the slope of the curve in the linear region. This voltage-temperature relationship is important to determine the noise equivalent temperature difference (NETD) of the sensor. Given an rms noise voltage at the output of the sensor, the NETD is

$$NETD = \frac{V_{noise}}{SITF} \quad [°\text{C or K}] \tag{8.21}$$

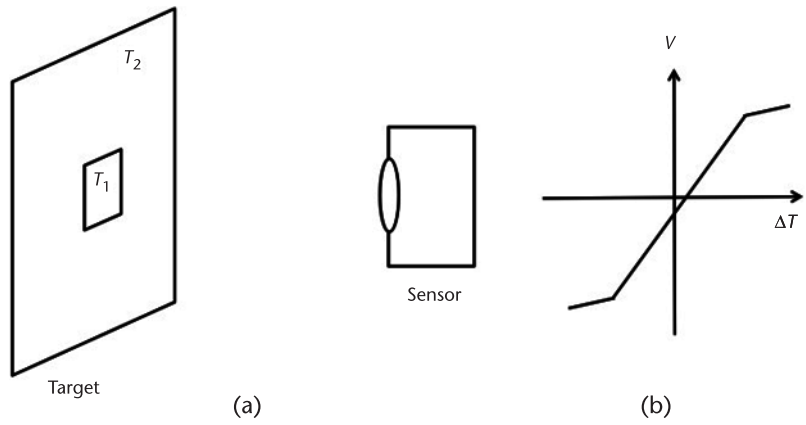


Figure 8.11 System intensity transfer function: (a) measurement and (b) curve.

The NETD is a sensitivity measure of infrared systems that describes noise in terms of source temperature variation. The definition and estimate of NETD are discussed later in this book.

8.6 Detector Sensitivity

The responsivity of a detector relates the signal level, in volts or amps, to the incident flux on the detector. A large responsivity does not mean that the detector can be used to easily discern small optical signals. An SNR is required to determine whether the signal can be differentiated from the noise. Because a large responsivity may not correspond to a large SNR, the term *sensitivity* was developed to describe a detector's SNR characteristics. The SNR at the output of a detector can be described [12] as

$$\frac{S}{N} = \frac{\Phi}{i_n} R \quad \text{[unitless]} \tag{8.22}$$

where R is the responsivity of the detector in amps per watt, Φ is the flux on the detector in watts, and i_n is the noise current in amps. If the SNR is set to 1 to determine a noise equivalent incident power (NEP), we can rewrite (8.22) as

$$NEP = \Phi_{S/N=1} = i_n/R \qquad \text{[W]} \tag{8.23}$$

If the responsivity is written in terms of volts per watt, then the NEP would be v_n/R. NEP is a function of wavelength and of the optical signal frequency (modulated or chopped light). That is, $R = R(\lambda, f)$ and NEP = $NEP(\lambda, f)$. NEP is also a useful quantity in that it provides a noise equivalent flux on the detector so that we can determine the overall system SNR. This is assuming that we know the noise current or the noise voltage on the output of the detector.

A better figure of merit is D^* ("dee star") or normalized detectivity. The D^* is also a function of wavelength and modulation frequency and can be written as follows:

$$D^*(\lambda, f) = \frac{\sqrt{A_d \Delta f}}{NEP} \quad [\text{cm-Hz}^{1/2}/\text{W}] \tag{8.24}$$

where A_d is the detector area in square centimeters and Δf is the noise equivalent bandwidth of the system that limits the detector output noise. When D^* is specified, it is normalized to a 1-cm^2 area and a 1-Hz noise equivalent bandwidth. With D^*, a larger value corresponds to a higher sensitivity. The unit of $\frac{\text{cm} - \text{Hz}^{1/2}}{\text{W}}$ is also known as a Jones. In addition, there is a blackbody D^*, where the radiant power on the detector is that of a blackbody, so the D^* is a broadband response that is a function of temperature and frequency $D^*(T, f)$.

For photon detectors, spectral D^* is an increasing function to a peak D^* near the cutoff frequency. Figure 8.12 shows the typical spectral D^*. The cutoff wavelength occurs when the incident wavelength does not exceed the material bandgap energy. The blackbody D^* is the equivalent sensitivity of the detector over all of the blackbody emitted wavelengths. The blackbody D^* is always less than the peak D^* because the detector does not respond to the blackbody wavelengths beyond the detector cutoff frequency.

One of the more common questions addressed by sensor designers and analysts involved with the use of detectors and SNR calculations is "Does my SNR increase or decrease with a larger or smaller detector?" Although this may seem like a simple question, the answer is, like many answers, "It depends." We consider the solution as an exercise in detection of resolved and unresolved sources.

Consider a simple sensor with a single lens and a single round detector as shown in Figure 8.13. The sensor is a large distance from the source, so that the image of the source is one focal length from the lens. Also, the atmospheric and optical transmissions are assumed to be 1 (i.e., there are no losses). Resolved sources (sometimes called extended sources) are those that more than fill the detector's instantaneous field of view (IFOV) as shown in Figure 8.13. Because the target is larger than the detector's FOV, not all of the target's emitted radiation that enters the sensor's input aperture falls on the detector. Any radiation leaving the source that is outside the sensor's FOV cannot fall on the detector.

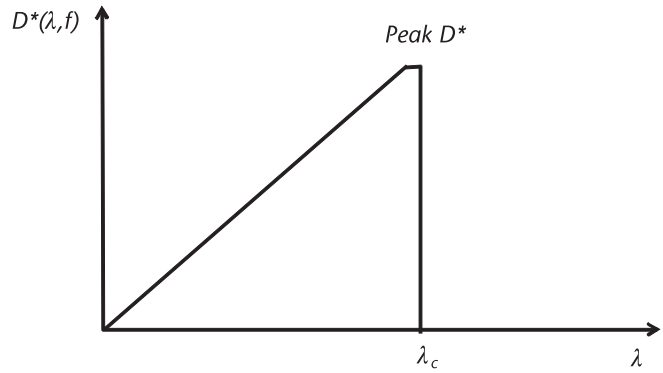


Figure 8.12 Spectral D^*.

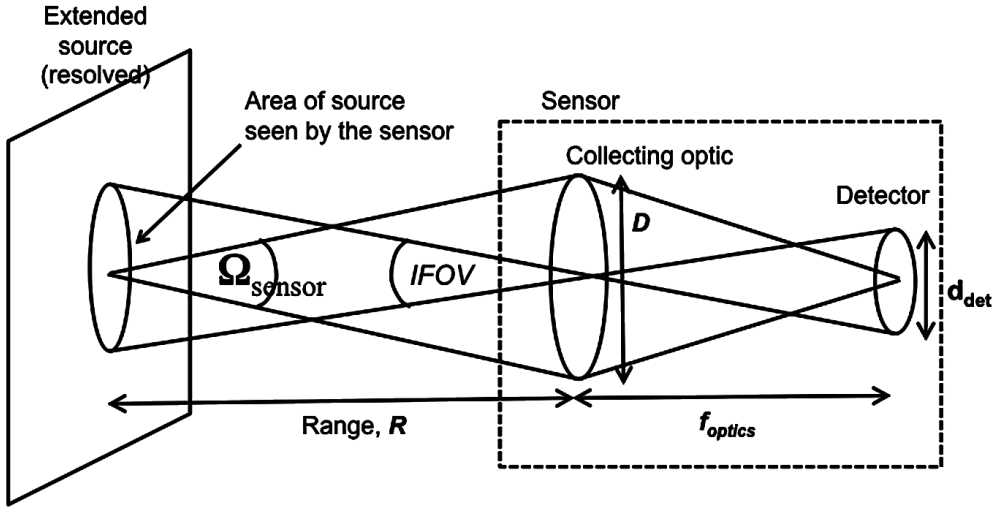


Figure 8.13 Resolved source.

The resolved-source example is one that is common in imaging systems because a large number of detectors (i.e., a detector array or scanned detector) are designed to image a single source. Therefore, the detectors may view only a small part of the entire target. The common problem seen in most applications for the resolved source is to determine the amount of the target flux impinging on the detector. We treat the sensor as a spectral instrument. The first source quantity considered, as usual, is the emittance of the source $M(T, \lambda)$. If the target is not a blackbody, then the emittance is given by $\varepsilon(\lambda)M(T, \lambda)$ where $M(T, \lambda)$ is the emittance of a blackbody at the temperature of the target. Assuming a Lambertian source, the radiance of the source is given by

$$L(T,\lambda) = \frac{\varepsilon(\lambda)M(T,\lambda)}{\pi} \quad [\text{W/cm}^2\text{-sr-}\mu\text{m}] \tag{8.25}$$

The radiance here applies to the entire source surface if the emissivity and temperature of the source are uniform over the source area. Only flux from within the source area corresponding to the sensor is seen by the detector. Therefore, an effective intensity (as seen by the sensor) can be given for the area viewed by the detector. The area of the source seen by the detector is

$$A_s = \frac{\pi R^2[IFOV]^2}{4} \quad \frac{\pi R^2[IFOV]^2}{4} \quad [\text{cm}^2] \tag{8.26}$$

where R is the range from the source to the sensor and is given in centimeters. Also, the small-angle approximation is used where the sine of the IFOV is equivalent to the angle itself when measured in radians. This is a reasonable approximation for angles of less than 5 deg.

The equivalent intensity of the area seen by the sensor can now be calculated. The intensity is given by the source area multiplied by the radiance:

$$I(T,\lambda) = L(T,\lambda)A_s = \frac{\varepsilon(\lambda)M(T,\lambda)R^2[IFOV]^2}{4} \quad [\text{W/sr-}\mu\text{m}] \tag{8.27}$$

This is only the intensity provided by the source area seen by the sensor. A sensor with a larger IFOV views a larger source area and, hence, sees a larger intensity. However, the intensity given radiates outward from the source over the entire hemisphere. Only the flux directed into the entrance aperture of the sensor is collected onto the detector. The intensity is converted into flux on the detector by multiplying the sensor's solid angle by the intensity. The sensor's solid angle (as seen by the source) is given by

$$\Omega_{sensor} = \frac{\pi D^2}{4R^2} \quad [\text{sr}] \tag{8.28}$$

The intensity for a flat source falls off with the cosine of the off-axis angle (see Chapter 5), so it is assumed that the diameter of the sensor's entrance pupil is small compared with the range. Now, the flux into the sensor's entrance pupil (and subsequently onto the detector) is

$$\Phi(T,\lambda) = I(T,\lambda)\Omega_{sensor} = \frac{\pi D^2 IFOV^2}{16}\varepsilon(\lambda)M(T,\lambda) \quad [\text{W}/\mu\text{m}] \tag{8.29}$$

This is an important equation because the power on the detector is not a function of range. With the assumptions given in this section, the equation shows that no change in flux is observed by moving the sensor closer to or farther from the target. The only significant changes in flux from geometric considerations are due to the diameter of the sensor's entrance aperture and the instrument's IFOV. For a rectangular FOV, the process is the same with the exception that the source area is calculated for rectangular area. The rest of the above procedure applies.

Broadband instruments collect flux over some band of wavelengths. To compute this broadband flux, (8.29) is integrated over wavelength:

$$\Phi = \frac{\pi D^2 IFOV^2}{16}\int_{\lambda_1}^{\lambda_2} \varepsilon(\lambda)M(T,\lambda)\, d\lambda \quad [\text{W}] \tag{8.30}$$

Because detector output voltage (or current) can ideally be considered directly proportional to flux on the detector, we can measure this quantity directly. The output voltage of a detector is determined by placing the detector responsivity inside the integral of (8.30).

The NEP for the detector is estimated by

$$NEP(\lambda) = \frac{\sqrt{A_d \Delta f}}{D^*(\lambda)} \quad [\text{W}] \tag{8.31}$$

where the detector frequency is operating in a flat response region. The SNR can now be determined for a waveband:

$$\frac{\Phi}{NEP} = \frac{\pi D^2 IFOV^2}{16\sqrt{A_d \Delta f}} \int_{\lambda_1}^{\lambda_2} \varepsilon(\lambda) M(T,\lambda) D^*(\lambda)\, d\lambda \quad \text{[unitless]} \tag{8.32}$$

Finally, we can simplify (8.32) by providing the detector area as $\pi(d_{det})^2/4$ and setting IFOV to d_{det}/f. In both cases, d_{det} is the linear diameter (in centimeters) of the detector. Now the SNR becomes

$$\frac{\Phi}{NEP} = \frac{\sqrt{\pi} D^2 d_{det}}{8 f^2 \sqrt{\Delta f}} \int_{\lambda_1}^{\lambda_2} \varepsilon(\lambda) M(T,\lambda) D^*(\lambda)\, d\lambda \quad \text{[unitless]} \tag{8.33}$$

A number of qualitative relationships are brought to the forefront by (8.33). Some general rules that describe the SNR of a detector viewing a resolved source are as follows:

- A small focal length and large optics diameter is desired (a small *f*-number). This relationship gives a squared (very strong) improvement in SNR.
- A larger detector diameter (or root of the area) gives a larger SNR, but poorer spatial resolution.
- A larger detectivity is desired (directly related).
- A smaller bandwidth is desired (related by the root). A small bandwidth means a longer photon integration time or longer dwell time.

Optimization of these parameters yields a system that can see very small changes in source exitance.

When the background photon flux is much larger than the signal flux and photon noise is the dominant noise source, we have the case *of background limited infrared photodetection* (BLIP). This is the case where the background (nonsource) light on the detector is larger than the signal flux. For terrestrial infrared systems, this case is common. The background objects including ground, ocean, and sky path radiance are that of 300K objects. These take up the majority of the FOV. The targets, or sources of interest, in the FOV are small and the temperatures (and corresponding emissions) are similar to that of the backgrounds such that the overall differential flux contributions are small. This situation corresponds to the unresolved source case.

We now consider the case where the source is unresolved (i.e., the source is smaller than the IFOV). For sources that do not fill the sensor's IFOV, the analysis is very similar to that of the previous analysis. The exception here is that the source emittance is only considered over the size of the source for the signal calculation. The D^* is given to be BLIP where the D^* is considered to be viewing objects that are around 300K. Most ground-based sensors can be considered BLIP as terrestrial objects are close to the temperature at which most D^* are specified. This is the same D^* that was given in the resolved case, assuming that the resolved source was at a typical terrestrial temperature. Consider the geometry of the unresolved target case in Figure 8.14.

Small sources are typically specified by an intensity parameter. To obtain a better understanding of the radiometric configuration, however, we start with the exitance of the source. If the source is Lambertian, the source radiance is

$$L(T,\lambda) = \frac{\varepsilon(\lambda) M(T,\lambda)}{\pi} \quad \text{[W/cm}^2\text{-sr-}\mu\text{m]} \tag{8.34}$$

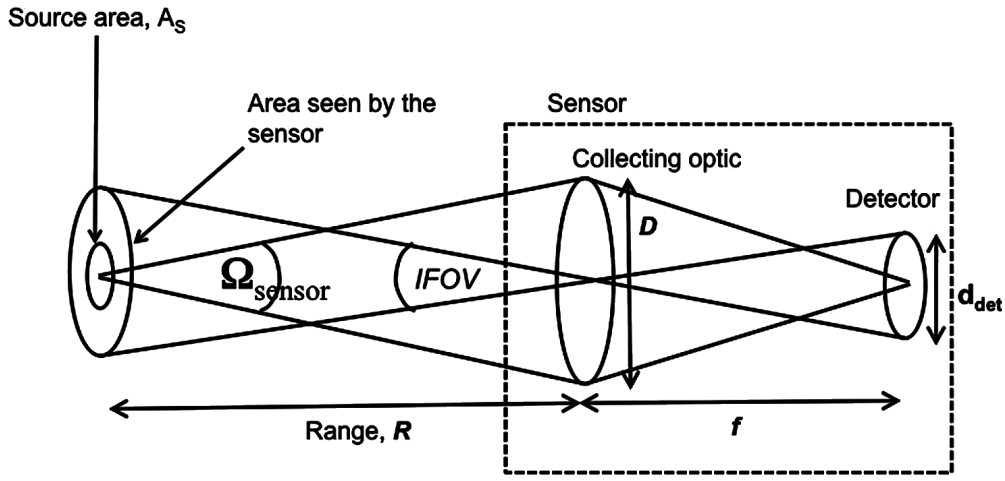


Figure 8.14 Unresolved source.

The intensity of the source is the radiance integrated over the source area. For a uniform radiance source:

$$I(T,\lambda) = L(T,\lambda)A_s = \frac{\varepsilon(\lambda)M(T,\lambda)A_s}{\pi} \quad [\text{W/sr-}\mu\text{m}] \tag{8.35}$$

To determine the amount of flux on the detector, the sensor's solid angle (as seen by the source) is required:

$$\Omega_s = \frac{\pi D^2}{4R^2} \quad [\text{sr}] \tag{8.36}$$

Again assume that the diameter of the sensor's entrance pupil is small compared with the range. The flux on the detector is now

$$\Phi(T,\lambda) = I(T,\lambda)\Omega_s = \frac{\varepsilon(\lambda)M(T,\lambda)A_s D^2}{4R^2} \quad [\text{W/}\mu\text{m}] \tag{8.37}$$

Note the flux dependence on range. The broadband flux is calculated by integrating the previous equation over wavelength:

$$\Phi(T,\lambda) = \frac{A_s D^2}{4R^2}\int_{\lambda_1}^{\lambda_2} \varepsilon(\lambda)M(T,\lambda)\,d\lambda \quad [\text{W}] \tag{8.38}$$

where, again, this value can be measured directly because the detector output voltage is proportional to input flux.

Using (8.31) for detector background-limited noise, the SNR becomes

$$\frac{\Phi}{NEP} = \frac{A_s D^2}{4R^2\sqrt{A_d \Delta f}}\int_{\lambda_1}^{\lambda_2} \varepsilon(\lambda)M(T,\lambda)D^*(\lambda)\,d\lambda \tag{8.39}$$

Where the source is smaller than the sensor FOV, the guidelines for SNR are as follows:

- A decrease in range gives a squared improvement.
- A larger optics diameter is desired and gives a squared improvement.
- A larger detectivity is desired and is related directly.
- A smaller detector is desired and gives a root improvement.
- A small bandwidth is desired and gives a root improvement.

A BLIP scenario assumes that the target signal is small compared with the background signal. The background signal limits the detection of the signal through the specification of D^*, where D^* is usually given for 300K backgrounds. If the background is cooler, D^* becomes higher, and if the background is hotter, D^* becomes lower. In some space applications where a small signal is discerned from backgrounds that are extremely cold, the D^* can be extremely high. If a detector is viewing a small target inside a furnace, where the background is high in temperature, then the D^* can be very low. BLIP conditions are usually satisfied in most tactical scenarios and many strategic scenarios. Images with the ground, ocean, and horizontal atmospheric backgrounds provide a large majority of BLIP conditions. However, space and vertical atmospheric backgrounds are cold and are characterized by different conditions.

The resolved SNR and the unresolved SNR can be compared directly if BLIP conditions are met. This would mean that the source temperature for the resolved case would be around 300K. For the unresolved case, the background is also near 300K. We also assume that the target signal is not too far from 300K, so that the D^* does not vary significantly. It is worthwhile to evaluate the SNR as a function of detector size to determine the optimal SNR configuration of a sensor. We provide this comparison through the following example.

Example 8.4

A round object at 310K is positioned in front of a 300K background. The source is 10 m in diameter at a range of 100 m. A round detector is operating in a 9- to 10-μm band, where the D^* is considered to be a constant $1 \times 10^{10} \frac{cm - Hz^{1/2}}{W}$ for BLIP conditions. The D^* can be considered constant over this band and is approximately the same for the 310K and 300K conditions. The signal is chopped with blades at 300K so that no differential signal is seen when the detector views the background. Assume that the emissivities of the blades, background, and signal are 1 (i.e., no losses). The noise equivalent bandwidth is 1,000 Hz. The optical system has a 10-cm entrance aperture diameter and a 20-cm focal length. Assume that the optical transmissions of the atmosphere and the sensor are 1. Determine the optimal round detector size for the largest SNR along with an estimate of the SNR.

The solution is provided by using (8.39) for the unresolved case with band-averaged values for D^*, emissivity, and differential emittance:

$$\frac{\Phi}{NEP} = \frac{A_s D^2}{4R^2\sqrt{A_d \Delta f}} \varepsilon \Delta M D^*$$

Equation (8.33) is used for the resolved case with the same band-averaged assumptions:

$$\frac{\Phi}{NEP} = \frac{\sqrt{\pi} D^2 d_{\text{det}}}{8 f_{optics}^2 \sqrt{\Delta f}} \varepsilon \Delta M D^*$$

The differential emittance can be found using the blackbody equation within the 9- to 10-μm band for a source of 300K and a source of 310K. The corresponding emittances are 3.119 and 3.672 mW/cm^2, respectively, giving a differential emittance of 0.553 mW/cm^2. If all of the above parameters are placed in the above equations and the detector size is varied, the resulting plot is shown in Figure 8.15.

With a small detector size, the source is resolved so that the signal grows with an increase in detector diameter. Once the detector diameter reaches the angular subtense of the source, the SNR is at a peak. The source subtends an angle of 5.7 deg that is matched by a 2-cm detector diameter at a focal length of 20 cm. At this detector diameter, the SNR is just under 20,000. As the detector size becomes larger, the source becomes unresolved. The source power decreases with the source area where the noise decreases with the root of the source area, so the SNR decreases. An important SNR lesson in this example is that a detector's SNR is at a maximum when the detector is matched to view the source area.

8.7 Detector Angular Subtense

While the SNR is maximum when a detector is matched to a source, a single detector sample of a source does not provide an image of the source. That is, a single radiometric sample of the source exitance cannot be mapped into a reasonable spatial depiction of a target. A large number of detector samples is desired across the

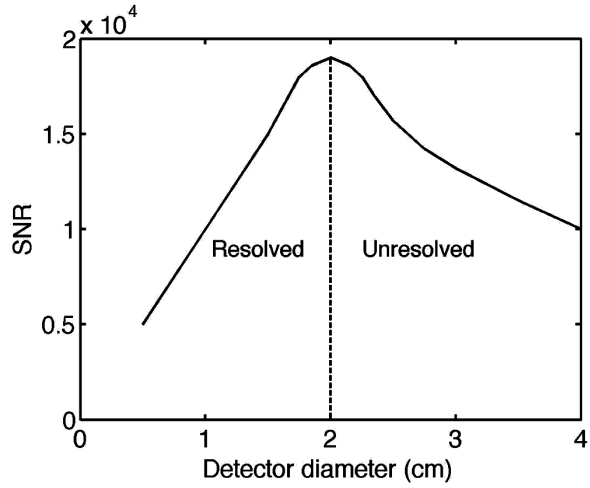


Figure 8.15 SNR as a function of detector diameter.

source so that a high-resolution picture can be constructed of the source exitance. The detector angular subtense (DAS) was developed to describe the resolution limitations of the detector size. For a square detector, there are two DASs: a horizontal DAS and a vertical DAS. Figure 8.16 shows that the DAS is the detector width or height divided by the focal length.

For the horizontal DAS of a square detector, the DAS is the detector width divided by the effective focal length of the optics:

$$\alpha = \frac{a}{f} \quad \text{[rads]} \tag{8.40}$$

and the vertical DAS is the detector height divided by the effective focal length:

$$\beta = \frac{b}{f} \quad \text{[rads]} \tag{8.41}$$

The horizontal and vertical DASs are usually specified in milliradians; the quantities described in (8.40) and (8.41) must therefore be multiplied by 1,000. The DAS describes the very best resolution that can be achieved by an EO or I^2R system due to the detector size limitations. Two point sources of light separated by 0.1 mrad cannot be resolved as two sources if a system has a focal plane array with DASs of 0.25 by 0.25 mrad.

Example 8.5

Two stars are separated by 2 mrad and are imaged by an EO sensor with a CCD detector array. The CCD array has rectangular detectors that are 35 μm in width and 50 μm in height with 100% fill factor (this means that the detectors are placed side by side with no gaps between detectors). The sensor has a focal length of 20 cm. Given that the sensor resolution is detector limited (the optical blur diameter is smaller than the detector), determine whether the two stars can be discriminated. The solution is determined simply by taking the DAS and comparing it with the star separation. The horizontal DAS is $\alpha = (35 \times 10^{-6}\,\text{m}/20 \times 10^{-2}\,\text{m}) \times 1{,}000 = 0.175$ mrad and the vertical DAS is $\beta = (50 \times 10^{-6}\,\text{m}/20 \times 10^{-2}\,\text{m}) \times 1{,}000 = 0.25$ mrad. With a 100% fill factor, the images of the stars would be separated by around 11 detectors in the horizontal direction or around 8 detectors in the vertical direction. Note also that the DAS for a circular detector is described only by the detector diameter divided by the effective focal length.

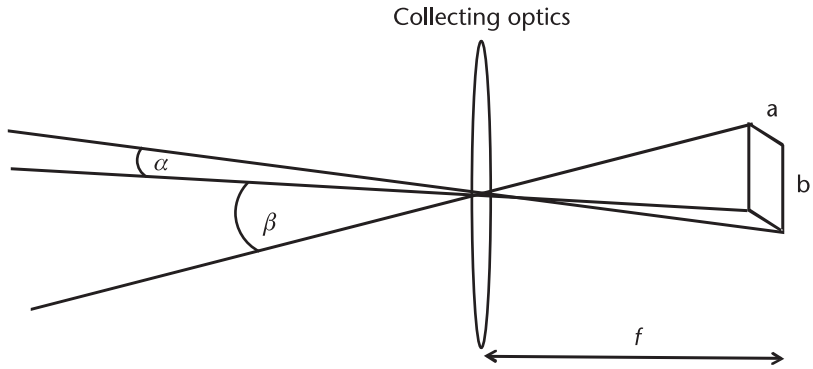


Figure 8.16 Detector angular subtenses.

The IFOV of a detector is sometimes confused with DAS, but Holst [10] describes the IFOV as the solid angle (in steradians) subtended by a detector. In the tradition of other solid angles, the IFOV would be defined as the detector area divided by the square of the effective focal length. It is quite common to see IFOV described in the same manner as the DAS, however, the IFOV is given as both DASs in milliradians. Both conventions are used throughout this book.

8.8 Scanning Configurations and Implementations

Although scanning systems are still being fielded as the primary sensor used by the U.S. Army, they are quickly becoming the minority in the other services. They will eventually become the minority sensor in the U.S. Army with the introduction of the third-generation FLIR, a staring dual-band MW/LW sensor. Scanning systems will continue to provide an advantage for some cases of large area coverage sensors. In these systems, multiple locations on the ground are shared with one detector (through scanning) where a good SNR is maintained. We cover scanning detectors here for completeness.

A scan mechanism is required for sensor systems that have either a single element or a linear detector array in order to move the image formed by the optics across the detectors. In this way, the detectors are distributed evenly over the image so that the image can be spatially represented. A scanner is not required for sensor systems that have large two-dimensional arrays. When infrared sensors were first developed and built in the 1970s and 1980s, the detector semiconductor manufacturing technology was not mature enough to build large-format arrays. Therefore, various scanning approaches were derived to allow either single element or linear array detectors to be scanned to form an image. Since that time, the detector technology has matured significantly. With respect to their detector array configurations, imaging sensor systems can be broadly grouped into two categories: *scanning* and *staring*. Scanning systems can be further subdivided into *serial* and *parallel* systems. Table 8.1 summarizes the array configurations and scanning types.

Staring systems, as the name implies, involve a system approach in which the image plane area is filled with detectors that have been distributed onto a single substrate. With this approach, no scanning is required, although there is usually some type of periodic shuttering for gain-and-level correction. Staring array systems offer the potential of having a much higher sensitivity than scanning systems because they have much longer integration times. In the SNR calculation of (8.33), the temporal frequency is replaced with $(1/2\tau_{int})$, where τ_{int} is the detector integration time, for these systems. The higher sensitivity does come at the price of lower resolution because the detector elements have spacing limitations in both the horizontal and vertical directions. Note that a relatively new technique known as microscanning helps to circumvent this problem. Microscanning is a technique, whereby the line-of-sight is stepped or dithered to increase the effective spatial sampling frequency.

Table 8.1 Detector/Scanning Configurations

Configuration	*Single Element*	*Linear Array*	*2D Array*
Serial scan	X	X	
Parallel scan		X	X
Staring			X

For serial-scan systems, a scan mechanism is required to move the scene over a linear array of one or more fixed detector elements. The elements are aligned horizontally and scanned sequentially in a 2D rectilinear raster pattern to form an image, as shown in Figure 8.17. This type of system requires a two-axis scanner to produce an image. One variation of the scanning pattern shown is to have a number of adjacent detectors in the active or in-scan direction to help increase the SNR. The number of adjacent detectors is defined as the number of time-delay-integration (TDI) detectors, N_{tdi}.

One line corresponds to the detector output for a single active scan across the image (shown in Figure 8.17). The line direction, or in-scan direction, which is the direction of highest speed (15,750 lines per second for the U.S. RS170 standard 525-line format), is usually the horizontal direction. The cross- or vertical-scan direction is slower (60 fields per second for U.S. RS170 standard 525-line format). Sometimes the detector-line rate and number of lines do not correspond to the RS170 television standard, so a special display or a reformatter is required to interpolate the line information. Many older systems use this approach. Newer systems use high-definition television (HDTV) format, however, these systems are usually associated with staring arrays.

The parallel-scan system consists of a linear array that is orthogonal to the cross-scan direction. This linear array has one or more columns of detectors, while the number of detectors in the vertical direction is sufficient to cover the desired FOV. Three primary types of parallel-scan configurations are shown in Figure 8.18.

Figure 8.18(a) shows a linear array of N detector elements. The image formed by the optics is scanned horizontally across the linear array during which the detectors are sampled. At the end of the first scan, a slight tilt of the scanner in the vertical direction is applied and the image is scanned across the detector array in the opposite direction. This provides coverage in the region of the image plane, which was not covered with the first scan. Each scan forms a field and the two fields are combined to form a frame that represents a complete image. The field rate for these types of systems is typically 60 Hz, which produces a frame rate of 30 Hz. This scan configuration is used in the common module (or first generation) FLIR, which was developed by the military. The second figure, Figure 8.18(b), is similar to Figure 8.18(a) with the addition of serial (horizontal-scan) TDI. TDI is when two or more detectors are sampled in the in-scan direction at the same positions, and the outputs of the detectors are added together. The noise adds in

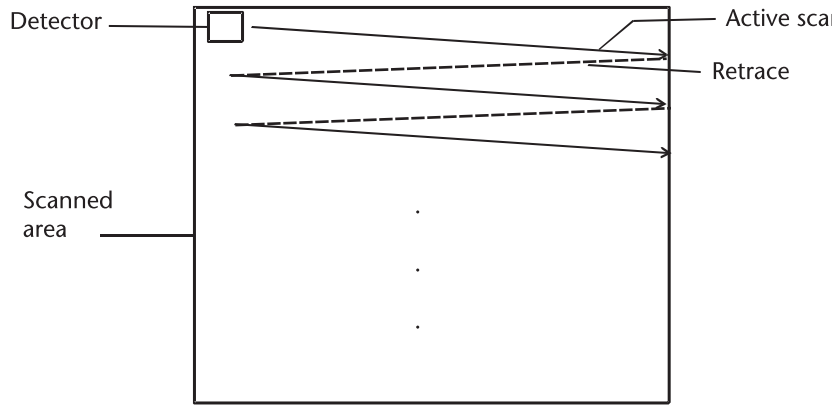


Figure 8.17 Two-dimensional serial (raster)-scan pattern.

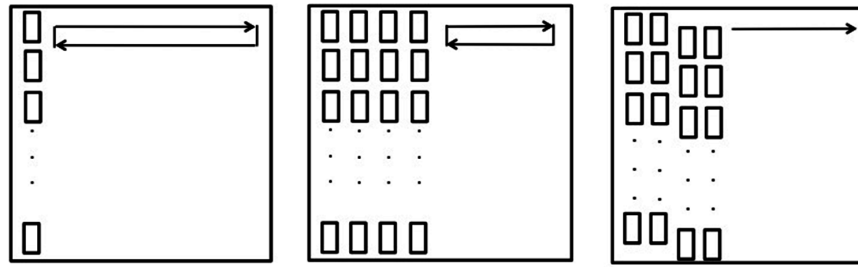


Figure 8.18 Parallel-scan configurations: (a) bidirectional interlaced scan, (b) bidirectional interlaced scan with TDI, and (c) unidirectional scan with TDI.

rms fashion and the signal adds directly giving a $\sqrt{N_{tdi}}$ improvement in SNR. The last configuration, Figure 8.18(c), includes a staggered array of detector elements along with TDI to allow the entire image plane to be covered with a single scan. In the detector array configuration, the increased number of detector elements in the vertical field of view (VFOV) gives redundant sample points. A second-generation FLIR is this type of system.

Several associated parameters are used to describe scanning-type sensor systems. The scanner efficiency, denoted by η_{sc}, is the ratio of the amount of time that the detectors actively sample the sensor's FOV during one frame time t_s for a given frame rate F_r:

$$\eta_{sc} = t_s F_r \tag{8.42}$$

where t_s is in seconds, and F_r is in frames per second. There are two primary reasons for the efficiencies being less than unity. The first is that the scanner traverses a distance larger than the horizontal field of view (i.e., dead space or radiometric reference scanning) and, second, the turnaround time of the scanner is significant because of the mass of the mirror. Typically, scanner efficiencies vary from as low as 50% to greater than 90%.

Two other parameters are the number of samples per horizontal DAS and the number of samples per vertical DAS. The number of samples per horizontal DAS is sometimes called the *serial scan ratio*, N_{ss}, and the number of samples per vertical DAS is sometimes called the overscan ratio N_{os}. There are other definitions for the serial-scan ratio [13]; however, we use the nomenclature just given. These parameters have been called by other terms, such as the number of samples per horizontal IFOV and the number of samples per vertical IFOV. Consider the detectors shown in Figure 8.19.

An N_{ss} of less than 1 shows that the detector sample is taken at positions that are farther apart than the horizontal DAS (or in linear dimensions, the separation of the detector samples is farther apart than the horizontal detector dimension in meters). The distance between samples in meters is sometimes called the *pixel pitch*. The sample spacing for N_{ss} equal to 1 gives a sample distance in radians that is equal to the horizontal DAS. Finally, the N_{ss} greater than 1 case gives a sample spacing that is less than the horizontal DAS, α.

The overscan ratio, N_{os}, of less than 1 means that the vertical sampling distance is farther apart than the vertical DAS. An N_{os} equal to one gives a vertical sample spacing equal to the vertical detector angular subtense, and an N_{ss} greater than 1 gives a vertical sample spacing smaller than the vertical DAS, β.

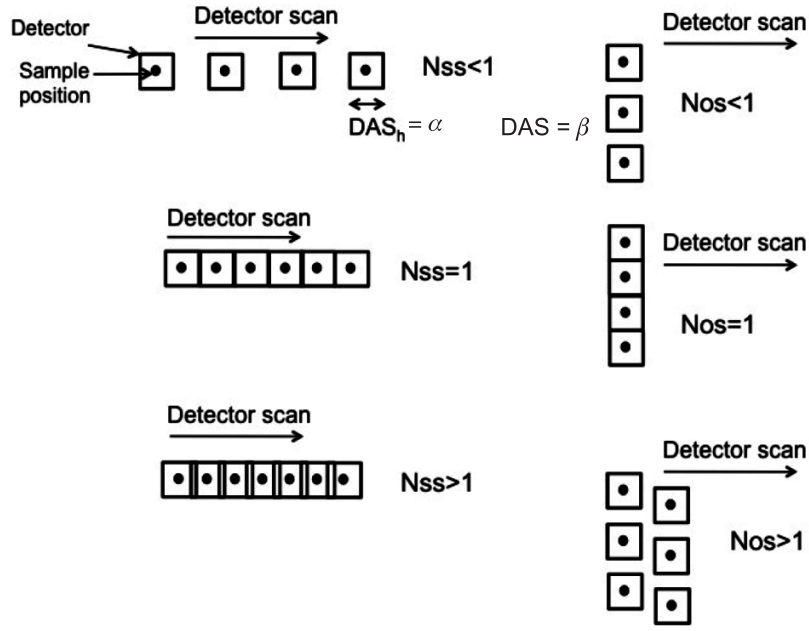


Figure 8.19 Serial-scan and overscan ratio.

The angular distance between samples is related to the serial-scan ratio and the overscan ratio along with the horizontal and vertical DASs. The horizontal angle between samples is given by

$$\delta_h = \frac{\alpha}{N_{ss}} \quad \text{[rad]} \tag{8.43}$$

and in the vertical direction

$$\delta_v = \frac{\beta}{N_{os}} \quad \text{[rad]} \tag{8.44}$$

For staring arrays, $\delta_h \leq \alpha$ and $\delta_v \leq \beta$, so that N_{ss} and N_{os} are both less than or equal to 1 unless microscanning is employed. A staring array where $\delta_h = \alpha$ and $\delta_v = \beta$ is considered a "100% fill factor" detector array. Microscanning takes the staring array and jitters it in a small pattern as shown in Figure 8.20, so that N_{ss} and N_{os} are greater than 1.

The entire array traverses the microscan pattern shown to the right of the detector array in Figure 8.20. The detectors are sampled a number of times (typically four times) during this scan pattern, which lowers the effective sampling distance.

Another important parameter is the detector dwell time t_d. This is defined as the time required for a detector to scan over a single point. The detector dwell time is given by

$$t_d = \frac{\alpha}{v_s} \quad \text{[sec]} \tag{8.45}$$

where α is the in-scan detector angular subtense in milliradians and v_s is the angular scan velocity in milliradians per second. For staring arrays, the dwell time is sometimes approximated as $1/F_r$ if the scan efficiency is nearly 1 (the frame transition time is the only nondetection period). Also, the bandwidth of the SNR equations given in this chapter is typically matched to the detector dwell time, so that $\Delta f \approx 1/(2t_d)$.

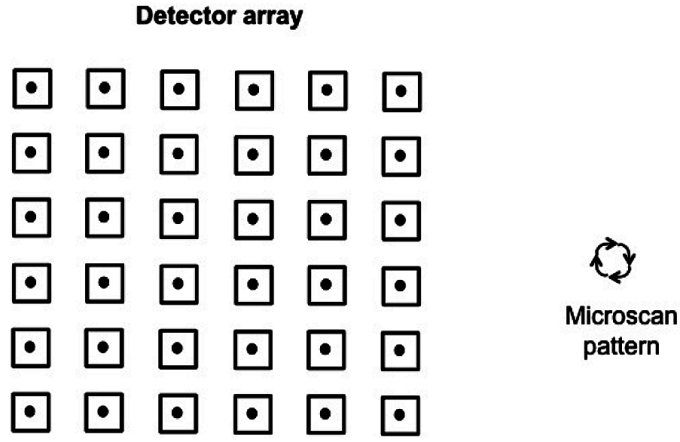


Figure 8.20 Microscanning.

A wide variety of methodologies (hardware) implement the scanning approaches discussed above. Scanning systems include mechanical, acousto-optical, and holographic scanning techniques. The most widely used are mechanical because they are the most efficient and well developed. There are two basic types of mechanical scanners: *parallel* and *converging beam* devices. These types differ in terms of where the scanner is placed in the optical path. The parallel beam scanner is placed in front of the final image-forming lens, and the converging beam scanner is placed between the final lens and the image. Of the mechanical scanners, the two most common are the rotating reflective drum and oscillating mirror or galvanometric scanners shown in Figure 8.21.

The rotating-drum scanner is ideal for systems that require a high constant velocity with which it can operate with very little loss in distortion of the imagery. These scanners are used primarily with parallel-beam systems because with a converging system, the oscillatory motion changes the focal point and hence causes severe distortion. The disadvantages of this scanner are cost and the inability to start and stop the scanner at a particular time and position. The galvanometric mirror oscillates periodically between two stops and can work with either parallel or convergent beams. This type of scanner is less expensive and more flexible than the drum scanners, but the oscillation becomes unstable near the field edges.

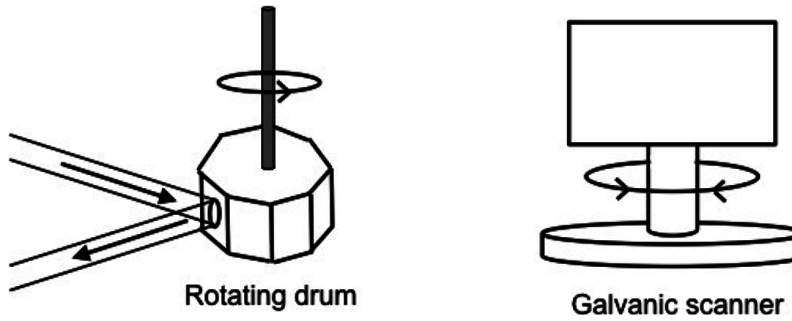


Figure 8.21 Common mechanical scanners.

8.9 Detector Transfer Functions

Whether a detector is scanned or staring, the detector integrates the incident flux over the spatial extent of the detector. That is, the output voltage of a detector can be determined by

$$V = \int_{\lambda_1}^{\lambda_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} i(\lambda, x, y) R(\lambda, x, y)\, dx\, dy\, d\lambda \quad [\text{V}] \tag{8.46}$$

where $i(\lambda, x, y)$ is the incident image irradiance on the detector corresponding to the spatial image and $R(\lambda, x, y)$ is the detector responsivity. From here, we are investigating the spatial transfer, so we drop the wavelength dependence where it is understood that we have the integrated flux for the output voltage. Consider the rectangular detector of Figure 8.22, centered at $[x', y']$. The detector integrates the incident image flux and the output voltage is sampled. A single-output value represents the image flux integrated by the detector:

$$o(x', y') = \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} i(x, y)\, rect\left(\frac{x - x'}{a}, \frac{y - y'}{b} \right) dx\, dy \tag{8.47}$$

resulting in a single voltage value located at $[x', y']$. Equation (8.47) is similar to a convolution where the output signal can be represented as

$$o(x', y') = i(x, y) ** rect\left(\frac{x}{a}, \frac{y}{b} \right) \Big| x = x', y = y' \tag{8.48}$$

where the substitution is performed only after the convolution is performed. The limitation on the single value can be accomplished by sampling. The evaluation at the single point is a single sample that can be obtained by the multiplication with $\delta(x - x', y - y')$ and then integrated (see the sifting property in Chapter 2).

For a staring array, scanning detector, or scanning linear array, the detector integrates the incident flux and, for now, we assume that the output is sampled at certain locations. Consider the detector sampling geometry of Figure 8.23.

The detector dimensions are a by b and are sampled at distances of c by d. The array size is limited by width w and height h. This geometry describes the majority

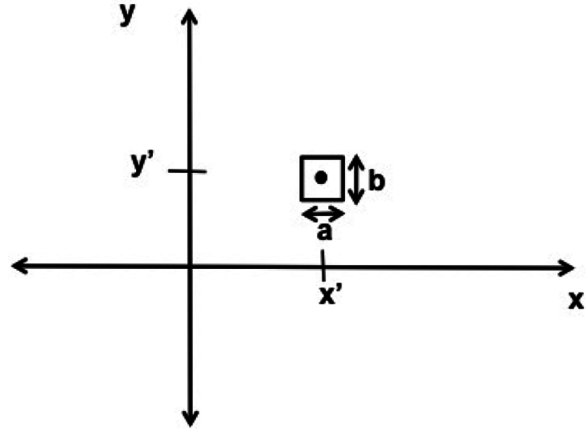


Figure 8.22 Detector spatial integration.

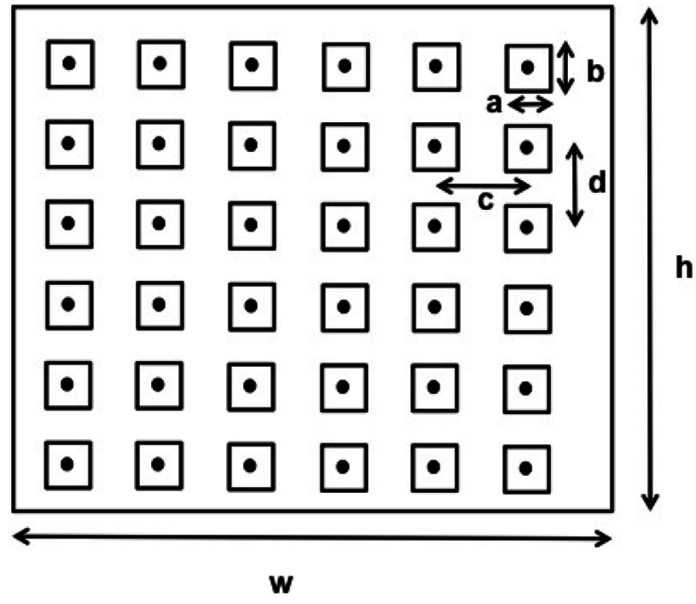


Figure 8.23 Detector sampling geometry.

of detector configurations including the scanned types mentioned in the previous section. The sampling distances c and d are simply a/N_{ss} and b/N_{os}, where the sampling distances can describe detector overlapping. The output from the sampled system is a set of discrete values that can be represented by

$$o(x,y) = \left\{\left[i(x,y) ** \frac{1}{ab}\, rect\left(\frac{x}{a},\frac{y}{b}\right)\right]\frac{1}{cd}\, comb\left(\frac{x}{c},\frac{y}{d}\right)\right\}\frac{1}{wh}\, rect\left(\frac{x}{w},\frac{y}{h}\right) \qquad (8.49)$$

To understand the image transfer in linear system terms, we take the Fourier transform of (8.49):

$$O(\xi,\eta) = \left\{\left[I(\xi,\eta)\; sinc(a\xi,b\eta)\right] ** comb(c\xi,d\eta)\right\} ** sinc(w\xi,h\eta) \qquad (8.50)$$

which describes many aspects of detector sampling and provides a good deal of guidance in system design. If we assume an infinite detector array (not limited by w and h), then the output spectrum is the image spectrum multiplied by a *sinc* function and convolved with a *comb*. Figure 8.24 shows a representation of the spectrum in the horizontal direction.

The *sinc* function has a width of $2/a$ to the first zeroes. It is generally assumed that the energy past the first zeroes is insignificant, but in reality it should be given consideration. If this function is convolved with the *comb* function as in Figure 8.24, overlapping occurs (i.e., aliasing) if $1/a$ is larger than $1/2c$. Also, the figure shows the case where $1/a$ is equal to $1/c$ (this is the case where the samples are spaced at exactly the detector width or 100% fill factor) and significant aliasing occurs. The cross-hatched region shows the folded detector spectrum that corresponds to high spatial frequencies being represented as aliased lower frequency signals. This case corresponds to a nonmicroscanned staring array with spacings equal to the detector widths. To avoid or control aliasing and spurious responses, three characteristics can be exploited.

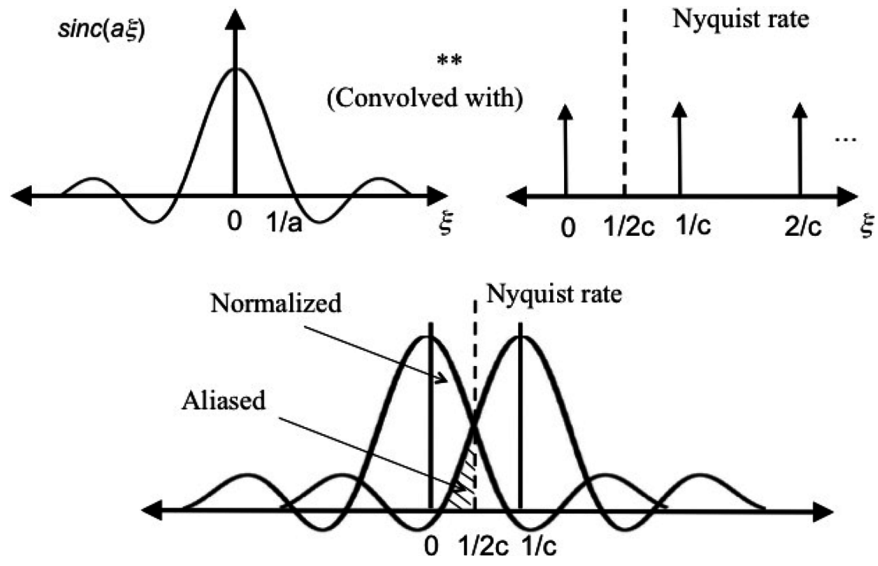


Figure 8.24 Detector output signal spectrum.

The first characteristic that we discuss is the least likely to be controllable. If the input scene (input to the sensor) has a spatial frequency band limit that ensures the highest spatial frequencies are less than $1/2c$ and $1/2d$ in cycles per meter (note that $c = N_{ss}a$ and $d = N_{os}b$ for scanned systems) or less than N_{ss}/α and N_{os}/β in cycles per radian, very little aliasing occurs. This requires the scene viewed by the sensor to have negligible spatial frequencies above these requirements. For high-altitude surveillance sensors, this is extremely unlikely. For tactical close engagements, it is more likely. Closer objects have a lower spatial frequency content.

One approach is to limit the input spatial frequencies by the optical design. This is to say that the optical MTF has cutoff frequency of d_{optics}/λ that matches N_{ss}/α or N_{os}/β in cycles per milliradian. Optical blur is usually more damaging than aliasing in system performance involving human vision. Systems with automated processing such as automatic target recognition or target tracking can be significantly impacted by aliasing. Finally, one approach is to oversample the detector output with at least two samples per DAS in both directions. This ensures that little aliasing and few spurious responses occur. In this situation, the limiting spatial frequencies are no longer N_{ss}/α and N_{os}/β, they are the first zeroes of the *sinc* function, $1/\alpha$ and $1/\beta$. This design approach includes an oversampled detector with an optical design that matches these spatial frequencies. Dithering or microscanning or super-resolution are ways to accomplish this level of sampling.

With any of the three above requirements met, the transfer function of the detector can be approximated by

$$H_{det}(\xi,\eta) = sinc(\alpha\xi, \beta\eta) \tag{8.51}$$

where the DASs are given in milliradians and ξ and η are given in cycles per milliradian. In using (8.51), be careful to meet aliasing requirements. This equation is commonly used with systems that inherently have significant aliasing, thus providing poor results in the estimation of system performance.

It is difficult to ensure that nonaliasing requirements are met. For infrared staring arrays, and even for second-generation FLIRs, oversampling by a factor of 2 cannot be achieved so the sensor design results in limited aliasing. The reduction in input spectrum by the optics effectively reduces the resolution in the output image. Some aliasing can be tolerated and image quality is a trade-off between resolution and aliasing.

Finally, we did not consider the last function in (8.50)—the window function, which describes the number of detector samples in the spatial array. Note that for a large array, the *sinc* function width is small. The convolution of any function with a function of small width decreases the resolution of the function by only a small amount. Remember that convolution with the zero-width delta function yields the exact function. A staring array of 1,000 × 1,000 detectors, where the detector spacing is very close to the detector width, produces a sinc function 1/1,000th the width of the detector sinc. The convolution of the output spectrum with the window sinc results in very little resolution effect. With an array of four detectors by four detectors, the convolution could affect the resolution significantly.

Example 8.6

A long-wave (assume a 10-μm wavelength) infrared sensor has an entrance aperture diameter (entrance pupil) of 6 cm and a horizontal detector angular subtense of 0.1667 mrad. The horizontal sampling frequency is 6 samples per milliradian. If the system images a point source, determine the percent of horizontal aliasing of the signal.

The solution can be determined by representing the point source as a *delta* function. Therefore, the input is $i(x,y) = \delta(x,y)$. This is a separable function so that $i(x) = \delta(x)$. The spectrum of the input is determined by the Fourier transform of the spatial input, so that $I(\xi) = 1$. That is, the spectrum is uniform over all frequencies. The detector output signal is

$$o(x) = \left[i(x) * h_{optics}(x) * \frac{1}{\alpha} rect\left(\frac{x}{\alpha}\right) \right] comb\left(\frac{x}{\delta_x}\right)$$

where h_{optics} is the optics impulse response, $\frac{1}{\alpha} rect\left(\frac{x}{\alpha}\right)$ is the detector impulse response, α is the horizontal DAS, and δ_x is the angular distance between samples. The only limit to the output spectrum of the detector is the optics MTF and the detector MTF, before detector sampling. In the frequency domain, the output spectrum is

$$O(\xi) = \left[I(\xi) MTF_{optics}(\xi) MTF_{det}(\xi) \right] * \delta_x \; comb(\delta_x \xi)$$

The optics MTF is assumed to be diffraction limited and is plotted in Figure 8.25. The detector MTF is plotted in the same figure for the given detector angular subtense. The optics and detector MTFs are multiplied before the sampling function. The sampling is characterized by the *comb* function, where the comb spikes are located at the origin and then every 6 cycles per milliradian apart. If the detector output spectrum is greater than 3 cycles per milliradian apart (this is the spatial Nyquist rate associated with the system), the convolution of the detector output spectrum results in an aliased signal. The amount of aliased signal, for the image of the point source, is the shaded area shown. This area divided by the total area of the output spectrum is calculated to be around 9%.

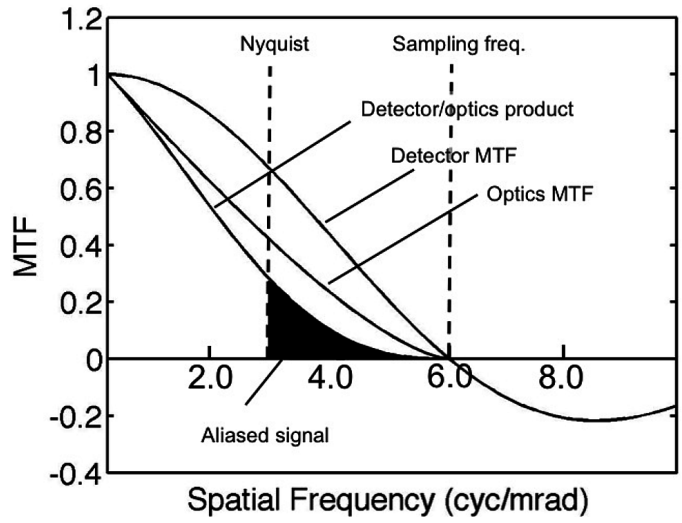


Figure 8.25 Example sensor.

We assume here that the atmospheric MTF is negligible. Because the sampling occurs after the optics MTF and the detector MTF (i.e., the sensor front end), the signal is aliased at this point in the system MTF cascade. The optical MTF, the detector MTF, and the sample spacing can be adjusted to reduce or eliminate aliasing. This is a common practice in sensor design, where the optics, detector, and sampling characteristics are matched to reduce aliasing while providing good image transfer. However, in many sensors, such as MWIR systems, aliasing is tolerated. The electronics, display, and human perception MTFs are performed after the scene is sampled and cannot eliminate the "in-band" aliasing.

8.10 Infrared Detectors

Research into infrared detection began more than 50 years ago before World War II, although few resources were devoted to this area because a significant amount of research funds was devoted to radar technology. While research continued over the years, it was not until the 1960s that the first FLIR sensors were developed and fielded. These earliest FLIRs were serial scanners. The technology continued to mature in the 1970s and 1980s, which resulted in the common module or first-generation FLIR. The sensors were based on linear mercury cadmium telluride (HgCdTe), often referred to as MCT, detector arrays that contained 60, 120, or 180 elements. These detector arrays were characterized by low quantum efficiency, limited dynamic range and sensitivity, single-band operation (8 to 12 μm), and were sensor noise limited. Even though second-generation FLIRs with scanned TDI detectors continue to be fielded, staring arrays have become the dominant detector in systems. The following paragraphs describe the current status of infrared detector materials.

MCT is the highest performance infrared material currently available. It has the greatest investment and the largest production base for infrared materials. Setting the ratio of mercury to cadmium in the detector growth process allows the waveband to be ad-

justed from near infrared (1 to 2 μm) to long-wave infrared (8 to 12 μm or LWIR). MCT is typically cooled to 77K for operation in the MWIR (3 to 5 μm) and LWIR regions. Current research and development in MCT is aimed at raising the operating temperature while keeping dark current low for MWIR applications. MCT is also the primary material used for dual-band or multiband operation (covered in a later section). LWIR and MWIR detection on the same detector focal plane is sometimes called third-generation FLIR and can be fabricated using vapor phase or advanced liquid phase epitaxy. A primary disadvantage of MCT is that the material is extremely difficult to produce without defects, which limits production yields and increases production costs (the impact is more dramatic on LWIR applications). Another disadvantage is that MCT requires either specialized or expensive substrates such as CdTe. Growth on alternate substrates such as silicon or GaAs has been successful but limited to MWIR applications.

InSb or "Insbee" is a high-performance material that operates only in the MWIR. InSb came along in the early 1990s due to substantial investment from the aviation community. InSb is also operated at 77K and has high efficiency, low noise, and high operability (small number of bad or noisy detectors). InSb has led the push for realizing large-format detectors (discussed later in this chapter). InSb has a cutoff wavelength of 5 μm and cannot be used for LWIR applications. Also, InSb material does not lend itself to dual-band or multiband detector fabrication.

A detector that is limited in production and applications is the quantum well infrared photoconductor (QWIP). These are grown from alternating layers of gallium arsenide compounds such as GaAs, AlGaAs, or InGaAs. Because QWIPs are grown using semiconductor materials with large commercial applications, the detectors can be grown in commercial foundries. They are grown using vapor phase epitaxy in alternating layers and it is possible to tune the waveband by alternating the materials and varying the layer thicknesses. QWIPS have historically had poor quantum efficiency (sometimes called conversion efficiency) and lower operating temperatures (65K). Because the LWIR region has plenty of photons, the primary application of QWIP detectors is in the LWIR. QWIPS have been fabricated for multiband, MWIR, and LWIR, but these detectors have limited performance.

Type II strained layer superlattice (TII SLS) detectors have been proposed as high-performance photon detectors that can operate at higher temperatures than MCT and InSb. TII SLS has alternating layers of InAs and GaSb grown by molecular beam epitaxy (MBE). These semiconductor materials are also used in the commercial sector (like QWIPS), so a reduction in cost could be seen if these materials continue to improve. TII SLS has been shown to have high quantum efficiency, but currently these detectors suffer from very short minority carrier lifetimes so that the charge cannot be collected fast enough (before the charge is lost). These detectors have higher dark current so that their performance is currently not as good as that of MCT. Research and development continues on SLS to understand the lifetime issue and to improve the performance to near theoretical (background-limited performance). SLSs have been fabricated in the MWIR and in dual-band (MWIR/LWIR) regions.

nBn detectors have also recently been proposed as high-performance photon detectors that allow high operating temperatures and low dark currents. nBn is short notation that refers to an n-type semiconductor material, then an undoped bandgap material barrier (for B), and then a thin n-type semiconductor layer. The barrier blocks the flow of majority carriers while allowing the flow of minority

carriers, resulting in less noise and higher operating temperature with the same or better performance than non-nBn structures. nBn can be implemented in a number of different materials with different spectral responses. The fabrication cost is low due to the simplicity of the material fabrication, making the material appropriate for low-cost, large-format arrays. nBn detectors have been fabricated using InAs for MWIR detectors. The current problem with nBn materials is that they have a 4.2-μm wavelength cutoff, which limits MWIR performance. Ongoing research efforts are under way to increase the cutoff wavelength to 5.0 μm.

The premier material for uncooled infrared detectors is vanadium oxide (VOx). Uncooled detectors are usually operated in the LWIR band to take advantage of the high-level infrared signals. VOx detectors are very tiny microbolometers where the material is suspended above a silicon readout circuit. The thickness of the material and its thermal isolation allows for a microbolometer's good performance, although its performance is not as high as that of photon detectors. Disadvantages of uncooled detectors are the long thermal "time constants" (similar to detector integration time) and the required low *f*-number optics for reasonable SNR. The other uncooled detector material is amorphous silicon, which provides not quite as high a performance (due to higher 1/*f* noise) as VOx, but is used for commercial applications. VOx detectors are currently produced in $1{,}024 \times 768$ arrays with 17-μm pixel pitch.

8.11 Electro-Optical Systems

In the previous edition of this book, the CCD was the preferred detector for EO (ultraviolet through near infrared) applications up to a 1.1-μm wavelength. CCDs "shifted" charge through bins one stage at a time. When the CCD was used as a digital imager, the pixels were p-doped metal-oxide semiconductor (MOS) capacitors that were biased for image acquisition. The photon to electron transition occurred at the semiconductor–oxide interface. The CCD was used to read the charge out. The CCD replaced photoemissive tube devices. CCDs are still used in a number of applications where high-quality image data are required.

CCD detectors come in three different formats: full frame, frame transfer, and interline. A full-frame detector is one in which the image area is always active so a mechanical shutter must be added to eliminate smearing while the image is read out. With frame transfer, the photosensitive area is quickly transferred to an area on the chip with an opaque cover and then the opaque area can be read out slowly while the active area is exposed again. Interline transfer has an interlace of photosensitive lines of detectors with lines that are masked. The active lines are shifted to the masked lines quickly and then the masked lines are read out slowly. The fill factor of these systems is at most 50%. The quantum efficiency of CCD detectors is around 70%. Cooling reduces the dark current of CCDs so that they can be used in lower lighting conditions.

High-gain CCDs can be used in a number of very low light applications. An electron multiplied CCD or EMCCD (L3Vision or Impactron CCD) has gain registers inserted in the shift registers so that when the charge is read out, it is multiplied in gain stages. The gain is applied in shift stages in a similar manner to avalanche diodes and is gained through impact ionization. The gain is small (around a few percent) through each stage, but as the number of stages are high, the overall gain can be high.

Image intensified CCDs (IICCDs) have an image intensifier mounted in front of the CCD. The image intensifier is a photoemissive device that comprises a photocathode, a charge transfer plate, and a phosphorous screen. Light falling on the photocathode generates electrons that leave the cathode surface. The electrons traverse tiny holes in the charge transfer plate and as the electrons bounce off the sides of the tubes, the plate displaces more electrons in a photomultiplier manner. Very large gains are seen in the transfer plate tubes. The phosphor screen converts the electrons into photons, which are then absorbed by the CCD detectors. The IICCD can be shuttered with the control voltage to the cathode and can be used for range-gating applications.

Fossum [14] stated that CCDs were likely to be replaced in the long term by active pixel sensors (APS) and sure enough, they have been. APSs are defined as detector arrays that have at least one transistor within the pixel unit cell. APS sensors eliminate the need for perfect charge transfer. That is, amplification occurs before the transfer of charge. Currently, the most common EO detector is the APS, which is sometimes called the CMOS detector since it is made with the complementary MOS process. The APS or CMOS pixel includes a photodetector and an active amplifier. These detectors are used widely in the commercial sector and significant progress in array size and pixel pitch has been seen in cell phones, webcams, digital cameras, and so forth. CMOS detector are faster than CCDs, consume less power, and require much less sophisticated fabrication facilities. CMOS detectors are now more widely used than CCDs in applications from high-end photography to mobile phone cameras.

Cannon recently announced their 120-megapixel CMOS detector array, which is a 13,280- × 9,184-pixel chip with a 2.2-μm pixel pitch [15]. The array can be read out at 9.5 frames per second with an 8-bit dynamic range. A 10-bit version is under development. The company claimed that the detector array is currently the world's highest resolution array at its size. Meanwhile, Pentax produces a 40-megapixel detector array that is used in their 645D camera, which sells for around $10,000 [16]. Finally, the Apple iPhone 4S and the forthcoming iPhone 5 have a standard 8-megapixel camera that will be around for at least a few years [17].

The pace of technology today yields many surprising developments in extremely short periods. EO systems have an added development thrust (over infrared systems) in that they are used widely in commercial and consumer applications. An example is the digital photographic camera, where the market continues to grow and the price of the system continues to decrease. These and many other EO systems, will continue to improve at rapid rates.

8.12 Noise

Detector noise was treated in a general sense in this chapter, with detectivity providing the means to calculate the noise signal on the output of a single detector. Overall system noise includes photon noise (from both the signal and the background), fixed-pattern noise, generation-recombination noise, shot noise, Johnson noise, $1/f$ noise, and many other noise sources. Usually, one of these noise sources dominates the SNR of the system. The cause and impact of each of these are very different.

Photon noise is the fluctuation in detector output signal due to the discrete arrival of photons and the corresponding electron state transition. For scanning single- and

linear-detector infrared systems, the photon noise from the background flux is usually the performance-limiting noise source. These scanning systems account for a large majority of LWIR systems. However, a large number of MWIR systems are currently built using the staring array configuration. These systems are typically performance limited by fixed-pattern noise. Fixed-pattern noise is the spatial "fixed" noise caused by the nonuniformity in the detector (and electronics) responses over the array. Each of the detectors in an array has its own input flux to output voltage characteristic. These characteristics are caused by detector material variations, impurities, imperfections, electrical circuit tolerances, and so on. The result is a fixed, noisy pattern on the output of the sensor when viewing a uniform scene. In the infrared, the fixed-pattern noise is treated as an increase factor of the background-limited case. In the EO wavelengths, the fixed-pattern noise is treated as an rms electron count in the output of a CCD charge transfer. The other noises described above are those associated with readout electronics (generation-recombination noise, shot noise, Johnson noise, and 1/*f* noise) and are described in the electronics chapter.

Noise is treated differently in EO systems and infrared systems. In EO systems, the dominant noise sources are readout noise (i.e., Johnson, 1/*f* noise, and other electronic noises) and fixed-pattern noise. In infrared systems, the dominant noise source is the background photon noise (i.e., BLIP). Also, the approach to noise analysis in EO systems is in equivalent rms electrons in a charge packet. In the infrared, detectivity is used to determine an NETD and a minimum resolvable temperature difference. For fixed-pattern, noise-limited infrared systems, a 3D noise analysis is performed that modifies the NETD values. Because the noise treatments are different in EO and infrared systems, we reserve the treatments until later in the systems chapters.

8.13 Basic Background-Limited Infrared Photodetection

Detector noise was treated in a general sense in this chapter with detectors treated from a system point of view. Sometimes, a basic calculation of the SNR is useful where the flux on the detector is given in basic photon irradiance E_{qsig} in photons per square centimeter. The signal to noise (numerator and denominator, respectively) when the signal includes all of the photons on the detector is

$$\frac{S}{N} = \frac{\eta E_{qsig} A_d q}{\sqrt{2\eta E_{qsig} A_d q^2 \Delta f}} \tag{8.52}$$

where η is the detector quantum efficiency, A_d is the area of the detector, q is the charge on an electron, and Δf is the detector bandwidth. It is easy to simplify (8.52) for the case where the photons on the detector are primarily signal photons and the detector noise is dominated by the discrete absorption of photons. The use of this quantity is extended to background-limited infrared photodetection in [18]. In the BLIP case, common in the infrared wavelength, the signal rides on a huge background pedestal. That is, all of the photons are not useful in making an image. For example, the photons coming from a terrestrial background are from a 300K source and the photons coming from a target are from a 302K source. The signal photons can be considered the photons generated by a 302K to 300K differential or a 2K source, and the background photons are generated by

a 300K background source. An image made up of all the photons would look completely white with very little modulation. At any rate, the SNR of the BLIP case (assuming a large background photon flux compared to signal flux) is

$$\frac{S}{N} = \frac{\eta E_{qsig} A_d q}{\sqrt{2\eta E_{qbg} A_d q^2 \Delta f}} \tag{8.53}$$

where the background photon irradiance E_{qbg} is in photons per square centimeter. Both (8.52) and (8.53) are useful in determining basic signal to noise for a signal-limited system and a background-limited system. Both of these equations assume that all other noise sources in the detector are small compared to the noise generated by the photon absorption.

8.14 New Infrared Detector Arrays

Large-Format Arrays

In the past 5 years, there has been an explosion of infrared detector development for large-format detectors in the MWIR region. These detectors are used in wide-area persistent-surveillance sensor systems [19]. These systems are flown over cities and areas of operation and when an event happens (e.g., suicide bomber, vehicle bomb, or improvised explosive device event), then the frames taken of the area are analyzed to determine where these events originated. Target tracking is a function of these types of systems.

The night versions of detectors in these systems are mostly MWIR with little work performed in the LWIR. One detector used in such a system is a 4,000 × 4,000 (or 4K × 4K) array of InSb produced by L3 Corporation. This detector array is sometimes called the 16-megapixel InSb focal plane and is shown in Figure 8.26. The detector pitch is 15 μm. L3 is currently working on an 8K × 8K array (or 64-megapixel array) with 10-μm pitch. The frame rate of the array will be 16 frames per second or 16 Hz.

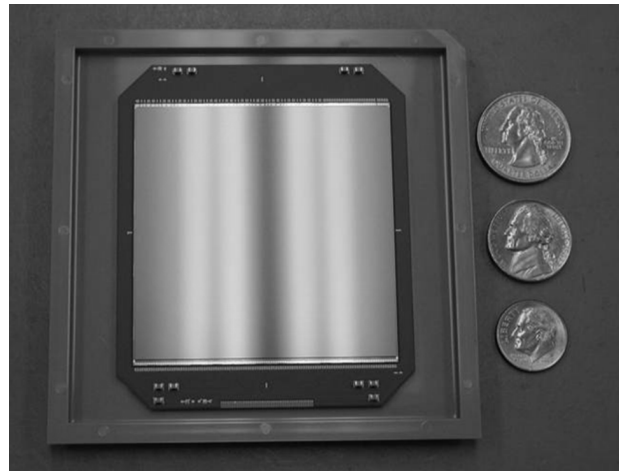


Figure 8.26 A 16-megapixel InSb focal plane by L3 Corporation. (Courtesy of Mark Greiner.)

There is also work by Raytheon on a large-format HgCdTe or MCT array that is 64 megapixels with a 10-μm pitch. This array would also be operated in the MWIR with an 8-Hz frame rate.

These large-format detector arrays are also being used for panoramic imaging with "donut" lenses to cover areas around a ship or a vehicle. For aviation applications, large linear arrays are swept across the FOV to accomplish large-area persistent surveillance with good frame rates; however, these linear arrays are used also used in the MWIR so must be configured in a TDI configuration with at least a few hundred detector elements in the cross-scan direction. Other large-format detectors under development range from staring, to scanned, to stare-stepped systems and the ones described here are good examples.

Dual-Band (Third-Generation FLIR) Detectors

Infrared dual-band systems are sometimes called third-generation FLIR systems and have both MWIR and LWIR detectors on a single focal plane. Oftentimes, the MWIR and LWIR detectors are stacked in a manner that achieves spatial registration. In the United Kingdom, dual-band (MW/LW) infrared detectors are HgCdTe grown by metal organic vapor phase epitaxy (MOVPE) [20]. This detector technology was developed by Selex Galileo Infrared Ltd. Their development led to a 640 × 512 dual-band array. This array size was fabricated with two detector pitches: 24- and 20-μm pitches. MWIR median NETD values achieved were 10 and 14 mK for the 24- and 20-μm pitch arrays, respectively. The corresponding LWIR median NETD values were 23 and 27 mK, respectively.

Raytheon developed dual-band detectors with a HgCdTe n-p-n triple-layer heterojunction (TLHJ) back-to-back diode detector structure [21]. The detectors (Figure 8.27) were grown using MBE on a cadmium-zinc-telluride substrate. The "bottom" n-type absorbing layer (band 1) had a shorter cutoff than the "top" n-type absorbing layer (band 2). Detector arrays were fabricated by etching a rectilinear pattern of narrow trenches in the TLHJ, which extended beyond the bottom p-n (band 1) junction. Only a single indium bump was then required to interconnect each physically and

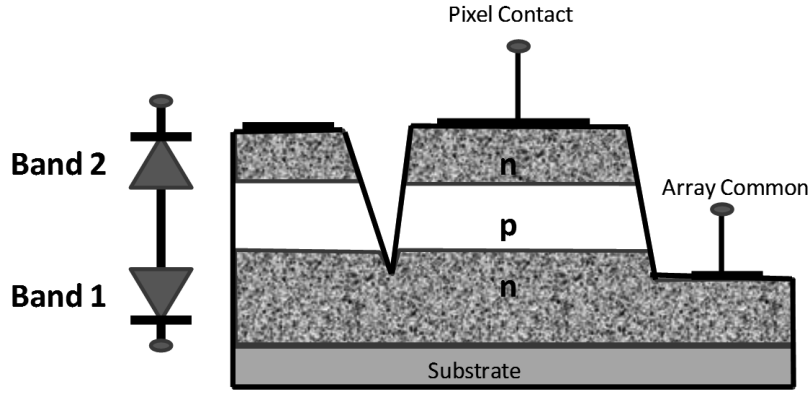


Figure 8.27 Raytheon dual-band detector structure. The lower (band 1) junction responds to shorter wavelength radiation (MWIR), while the upper junction (band 2) responds to longer wavelengths (LWIR).

electrically isolated back-to-back "mesa" diode to its corresponding unit cell input circuit in the readout integrated circuit (ROIC). The bias voltage polarity determined the infrared band that was read out. The resulting detector array was a 640 × 480 staring dual-band (MW/LWIR) focal plane. Raytheon Vision Systems has demonstrated dual-band MW/LWIR focal planes in 640 × 480 and 1280 × 720 formats with a 20-μm pitch.

A number of companies offer dual-band (MW/LW) focal planes in HgCdTe. SOFRADIR in France now produces a 640 × 512 array of 20-μm detectors. DRS, Teledyne, and AIM (Germany) also produce dual-band detectors. The current emphasis on dual-band detectors is to increase the array size, reduce the detector size, and increase the operating temperature all while maintaining good SNR.

Laser Range-Gated Detectors and Active/Passive Detectors

One imaging mode that has seen a great deal of progress in recent years for long-range identification is that of laser range-gated imaging. The U.S. Army investigated laser range-gated, shortwave IR (LRG-SWIR) imaging systems to perform long-range target identification [22]. The laser is used as a pulsed, flash-illuminator that is coupled to a range-gated, electron-bombarded CCD (EBCCD). Range-gated imaging of the laser pulse over a few nanoseconds provides a reduction in camera noise, selective image return, and a high level of anonymity over continuous-wave or longer pulsed lasers.

The target range must be known for range gating, so a laser rangefinder is used to determine the target distance prior to the image pulse. The laser is pulsed and a timer is initialized. After the appropriate pulse time of flight, the timer triggers the EBCCD electronic shutter at the return time of the pulse corresponding to the target range. The typical pulse duration is 10 ns, translating to a pulse length of roughly 3 m. Note that shuttering the camera at different trigger times can give different images of the target or the background. It is sometimes desirable to image past the target so as to give a target silhouette image. Gating the return signal also eliminates unwanted scatter from aerosols and debris in the atmosphere between the target and the sensor. Figure 8.28 shows the Army concept of operation where an inexpensive uncooled

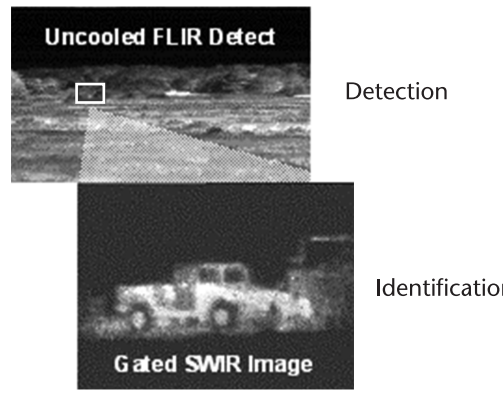


Figure 8.28 U.S. Army use of LRG-SWIR.

sensor provides the wide FOV search and detection function and the LRG-SWIR sensor provides the long-range target identification. LRG-SWIR is not good for wide field search and detection due to the laser power requirements of extended FOV. However, LRG-SWIR is great for high-resolution identification due to the shorter wavelength of SWIR compared to MWIR and LWIR.

The EBCCD was subsequently replaced by the Electron Bombarded Active Pixel Sensor (EBAPS, trademark of Intevac) detector array. The EBAPS has a GaAs photocathode with a high-resolution, backside-thinned, CMOS chip anode. The electrons emitted by the photocathode are directly injected in the electron-bombarded mode into the CMOS anode, where the electrons are collected, amplified, and read out to produce digital video directly out of the sensor [23]. Low noise gain is achieved by conversion of the high-energy photoelectron (1 to 2 keV resulting from the high-voltage bias applied across the photocathode and anode) to electron-hole pairs in the anode via the electron bombarded semiconductor (EBS) gain process. The EBS gain is a low-noise process. The current detector is a 1280×1024 chip with a 6.7-μm pixel pitch. These detectors work well not only with laser range-gated systems, but with low-light passive systems as well. The U.S. Air Force has already purchased 50 Litening G4 pods with the EBAPS detector and the laser range-gated imaging SWIR sensor [24]. It is likely that this detector technology will continue to proliferate.

An alternate to the EBAPS detector is the HgCdTe (or MCT) avalanche photodiode (MCT-APD). In the design shown in Figure 8.29 [25], there is a via in the middle of the detector with an n-type region surrounded by a p-type region of MCT. Figure 8.29 shows a cross-section side view of the detector. From the top view, the via is circular. The detector can be operated within the normal mode and responds to the MWIR passively. However, if the detector is reverse biased to avalanche mode, a low noise gain of over 100 is achieved. The ROIC shown is the ROIC where the charge capacity, sometimes called the well capacity, is 2.5×10^5 electrons. The detector has good quantum efficiency, low excess noise, high bandwidth, and a large dynamic range. The minimum integration time is 50 ns and the detector can

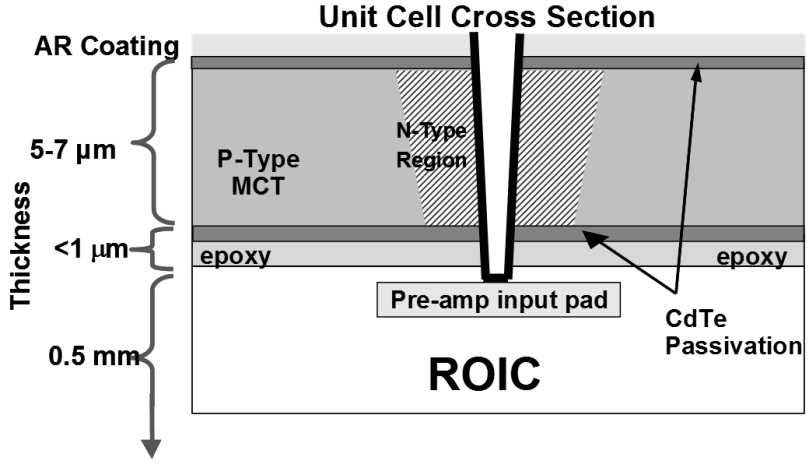


Figure 8.29 Side view of MCT-APD. (Developed by the U.S. Navy and DRS; courtesy of Jim Waterman, Navy Research Laboratory.)

operate at a 60-Hz frame rate. The operating temperature is typical of MWIR systems at 77K to reduce dark current. The current format is 640 × 480 with a 25-μm detector pitch and there are four vias per detector.

The current U.S. Navy program using the MCT-APD detector is called Active/Passive Dual-Mode Infrared Sensor (APDIS). The primary benefit of this detector over the EBAPS detector is that both passive MWIR sensing and LRG-SWIR sensing can be achieved using the same detector. The dual FOV optics must be part of such a system where MWIR is used for the wide field search and detection mode and the LRG-SWIR is used for long-range target identification. The U.S. Navy has been interested in SWIR for some time due to maritime haze penetration.

8.15 Exercises

8.1 Describe the differences between photon and thermal detectors and provide a list of applications for each type of detector.

8.2 Determine the bandgap energy of a detector material required to detect 10-μm monochromatic light. What occurs with a larger bandgap? A smaller bandgap?

8.3 A mercury-cadmium-telluride (HgCdTe) detector has intrinsic concentration such that n_n and n_p are both 1×10^{13} per cubic centimeter. The detector has a quantum efficiency of 60% once it is cooled to 77K. For a 50-mm detector depth, determine the change in voltage with respect to incident irradiance.

8.4 Determine the gain for a photomultiplier with six stages and a secondary emission ratio of 5.

8.5 Describe the difference between radiant and photopic responsivity. Why is photopic responsivity used and what are the risks associated with its use?

8.6 A round InSb detector is 1 cm in diameter with a perfect output filter of 1 kHz in bandwidth. Determine the noise equivalent power of the detector. State your assumptions.

8.7 Derive the SNR equations for the resolved and unresolved source cases where the detector is square and the source is square.

8.8 List all possible noise-reduction activities that apply to both resolved and unresolved source cases.

8.9 Calculate the vertical and horizontal detector angular subtenses for a 30- × 40-μm detector located at the image plane of a 25-cm focal length lens. Given that the detector is sufficiently sampled in both directions, how can the detector modulation transfer function be improved?

8.10 Explain why second-generation infrared sensors are BLIP limited compared to first-generation sensors.

8.11 For a bidirectional scanning sensor system that has a 60-Hz field rate (30-Hz frame rate) and 480 samples in the scan direction, determine the detector dwell time. (Assume that the scanner efficiency is unity.)

References

[1] Pinson, L., *Electro-Optics*, New York City, NY: Wiley, 1985, p. 103.

[2] *The Photonics Dictionary*, Pittsfield, MA: Laurin Publishing Company, 1996.

[3] Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, New York City, NY: Wiley, 1996, p. 86.

[4] Crowe, D., P. Norton, T. Limperts, and J. Mudar, *The Infrared and Electro-Optics Handbook, Volume 3: Detectors in the Electro-Optics Components*, Bellingham, WA: SPIE and ERIM, 1993, p. 177.

[5] Streetman, B., *Solid State Electronic Devices*, Englewood Cliffs, NJ: Prentice Hall, 1980, p. 65.

[6] Hergert, E., "Expanding Photodetector Choices Pose Challenges for Designers," *The 1996 Photonics Design and Applications Handbook*, Pittsfield, MA: Laurin Publishing Company, 1996, p.119.

[7] Grum, F., and R. Becherer, *Optical Radiation Measurements*, Volume 1, New York City, Academic Press, 1979.

[8] Finney, P., "IR Imaging With Uncooled Focal Plane Arrays," *Sensors, The Journal of Applied Sensing Technology*, 1996, pp. 36–52.

[9] Saleh, B., and M. Teich, *Fundamentals of Photonics*, New York City, NY: Wiley, p. 664, 1991.

[10] Holst, G., *CCD Arrays, Cameras, and Displays*, Winter Park, FL: JCD Publishing, 1996.

[11] Aikens, R., "Replacing Charge-Coupled Devices in Many Imaging Applications," *The Photonics Design and Applications Handbook*, Pittsfield, MA: Lauren Publishing, 1996.

[12] Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, New York City, NY: Wiley, 1996, p. 152.

[13] Shumaker, D., J. Wood, and C. Thacker, *FLIR Performance Handbook*, DCS Corporation, 1988.

[14] Fossum, E., "Active Pixel Sensors and Are CCDs Dinosaurs?" *SPIE Proceedings: CCD and Solid State Optical Sensors*, San Jose, CA, 1993.

[15] http://www.engadget.com/2010/08/24/canon-proudly-intros-120-megapixel-cmos-sensor-probably-wont-h/. Last accessed April 2012.

[16] http://www.engadget.com/2010/03/09/pentax-gets-official-with-40-megapixel-645d-medium-format-camera/. Last accessed April 2012.

[17] http://www.apple.com/iphone/features/#camera.

[18] Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, New York City, NY: Wiley, 1996, p. 212.

[19] Kruer, M. R., J. N. Lee, D. L. von Berg, J. G. Howard, and J. Edelberg, "System considerations of aerial infrared imaging for wide area persistent surveillance," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, G. C. Holst and K. A. Krapels (eds), Proc. of SPIE, Vol. 8014, 2011.

[20] Abbott, P., L. Pillans, P. Knowles, and R. K. McEwen, "Advances in Dual-Band IRFPAs made from HgCdTe grown by MOVPE," *Infrared Technology and Applications XXXVI*, B. F. Andresen, G. F. Fulop, and P. R. Norton (eds), *Proc. of SPIE*, Vol. 7660, 2010.

[21] King, D. F., J. S. Graham, A. M. Kennedy, R. N. Mullins, J. C. McQuitty, et al., "3rd-generation MW/LWIR sensor engine for advanced tactical systems," *Infrared Technology and Applications XXXIV*, B. F. Andresen, G. F. Fulop, and P. R. Norton (eds), Proc. of SPIE, Vol. 6940, 2008.

[22] Driggers, R. G., R. H. Vollmerhausen, N. M. Devitt, C. E. Halford, and K. J. Barnard, "Impact of speckle on laser range-gated shortwave infrared imaging system target identification performance," *Optical Engineering 42(03)*, 2003, p. 738–746.

[23] Aebi, V. W., K. A. Costello, P. W. Arcuni, P. Genis, and S. J. Gustafson, "EBAPS®: Next Generation, Low Power, Digital Night Vision," *OPTRO 2005 International Symposium*, Paris, France, 2005.

[24] http://defense-update.com/20111022_litening_g4_atp-se.html. Last accessed April 2012.

[25] Waterman, J., "Active Infrared Imaging Technology for Maritime Applications," *Infrared Technology and Applications XXXVII*, B. F. Andresen, G. F. Fulop, and Paul R. Norton (eds), *Proc. of SPIE, 2011*.

CHAPTER 9

Electronics

The function of electronics is to transform the detector output into a signal that can be viewed or processed (Figure 9.1). This transformation must be accomplished with minimal degradation of system performance. Issues of primary importance are low noise, high gain, low output impedance, large dynamic range, and good linearity. Because entire books have been written on sensor electronics, here we merely provide an introduction along with simplistic models that will permit the inclusion of electronics in a system performance estimate. For completeness, some of the scanning electronics are included from the previous version of this book. Many systems are still being fielded that are scanned even though they are being phased out in favor of staring arrays. For staring systems, we added a section for readout integrated circuits, or ROICs.

9.1 Detector Circuits

A large number of detector circuits are used to provide a voltage or current that is proportional to the irradiance on a detector. Bipolar, JFET, and MOSFET technologies are all used in detector amplification. We focus here on operational amplifiers (op-amps) because they are extremely common in detector circuit design. The circuits given here are useful as amplifiers for photovoltaic, photoconductive, bolometer, and pyrometer detectors.

For photovoltaic detectors, two types of circuits can be used as detector preamplifiers: the *voltage-mode circuit* and the *current-mode circuit*. The photovoltaic detector can be modeled with a voltage-current relationship that is very similar to that of a diode. The difference between the light-sensitive diode and the typical

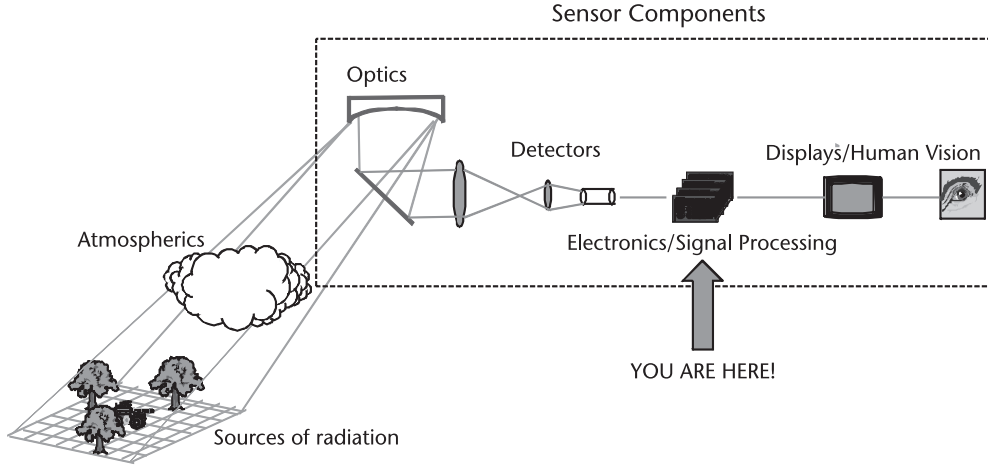


Figure 9.1 System electronics component.

diode is that the voltage-current relationship shifts with detector illumination. The magnitude of the voltage is related to the number of incident photons on the detector. This voltage-current relationship is the nonlinear diode equation

$$I = I_s[e^{eV/\beta kT} - 1] \quad [\text{A}] \tag{9.1}$$

where

I_s = detector saturation current in amps
e = charge on an electron (1.6×10^{-19} C)
β = efficiency factor (1 for ideal × 1 and 2 or 3 for the real case)
k = Boltzmann's constant (1.38×10^{-23} J/K)
T = detector temperature in Kelvin.

Whereas voltage is proportional to the number of incident photons, current is not linear with the number of photons.

Consider the voltage-current relationship and the voltage-mode circuit shown in Figure 9.2. The circuit is the standard noninverting op-amp circuit where the input current is negligible (recall that one of the ideal op-amp characteristics is that no current enters the op-amp). When the current is held to zero, there is a voltage change when the detector is illuminated. (This is illustrated in the voltage-current relationship of Figure 9.2.)

The ideal op-amp has an infinite open-loop gain. When the op-amp gain is sufficiently large (i.e., much larger than the closed-loop gain), the gain of the circuit can be approximated by the closed-loop gain equal to

$$A = \frac{V_o}{V_{det}} = \frac{Z_1 + Z_2}{Z_1} \quad [\text{unitless}] \tag{9.2}$$

where Z_1 and Z_2 are the impedances shown in the blocks of Figure 9.2. For an op-amp with an open-loop gain that is not much larger than the closed-loop gain, the voltage gain of the circuit is

$$A_v = \frac{V_o}{V_{det}} = \frac{Z_1 + Z_2}{Z_1 + \dfrac{Z_1 + Z_2}{A_{OL}}} \quad [\text{unitless}] \tag{9.3}$$

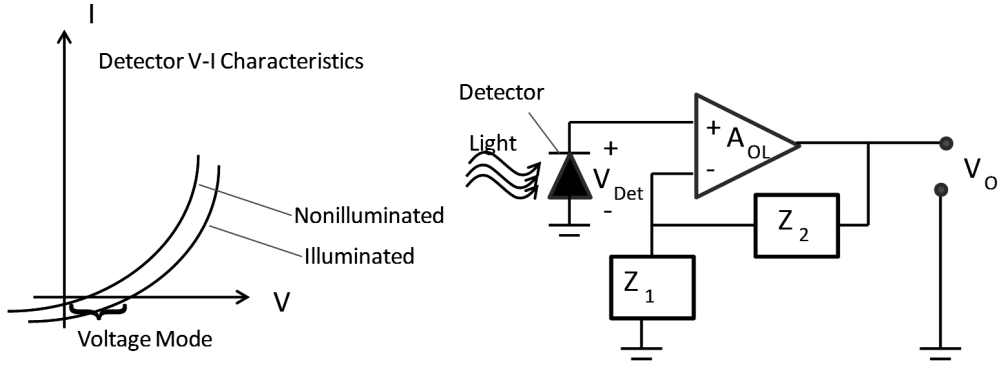


Figure 9.2 Voltage-mode detector circuit.

and A_{OL} is the op-amp open-loop gain. The output voltage of this circuit is in phase with the input voltage. Also, the voltage-mode circuit is desirable when a high-input impedance and a low-output impedance is desired.

The second type of photovoltaic detector circuit is the current-mode detector circuit show n in Figure 9.3. Recall that another ideal op-amp characteristic is that the voltage across the input terminals of the op-amp is essentially negligible (i.e., grounded in the configuration shown). Forcing the voltage to approach zero causes the detector to establish a current when the detector is illuminated as pointed out on the detector voltage-current relationship graph. The input impedance of the detector is not used in the determination of the gain. For an ideal op-amp, where the open-loop gain is large compared with the closed-loop gain, the amplifier voltage gain is

$$A_v = \frac{V_o}{V_{det}} = \frac{-Z_2}{Z_1} \quad \text{[unitless]} \tag{9.4}$$

The gain for a nonideal op-amp with a lower open-loop gain is

$$A_v = \frac{V_o}{V_{det}} = \frac{-Z_2}{Z_1} \frac{A_{OL}Z_1}{(1+A_{OL})Z_1 + Z_2} \quad \text{[unitless]} \tag{9.5}$$

This type of circuit is used for high-gain, nominal-input impedance, and low-output impedance. However, there is a signal inversion (i.e., the signals are 180 deg out of phase) from the detector voltage to the output voltage.

Example 9.1

Determine the voltage gain for the detector amplifier shown in Figure 9.4. Assume that the operational amplifier is an ideal amplifier (i.e., large open-loop gain). The solution is given as the current mode gain of (9.4), where Z_1 is the impedance of R_1 and C_1 in parallel. The impedance Z_2 is R_2 and C_2 in parallel. The impedance of a resistor and capacitor in parallel is

$$Z_1 = \frac{R_1}{1 + j\omega R_1 C_1}$$

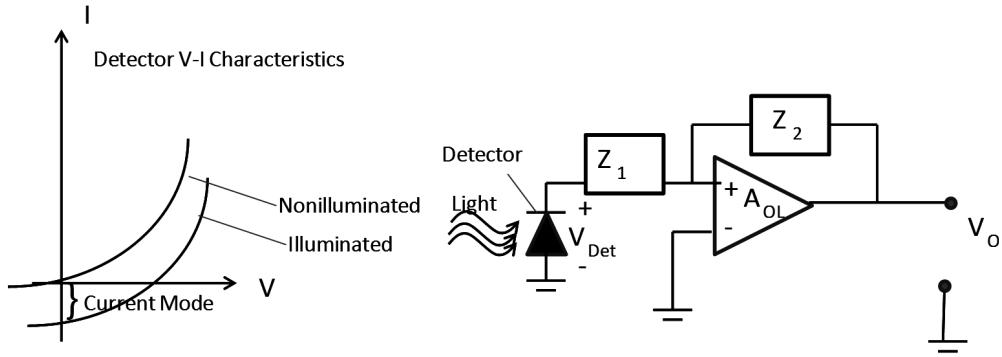


Figure 9.3 Current-mode detector circuit.

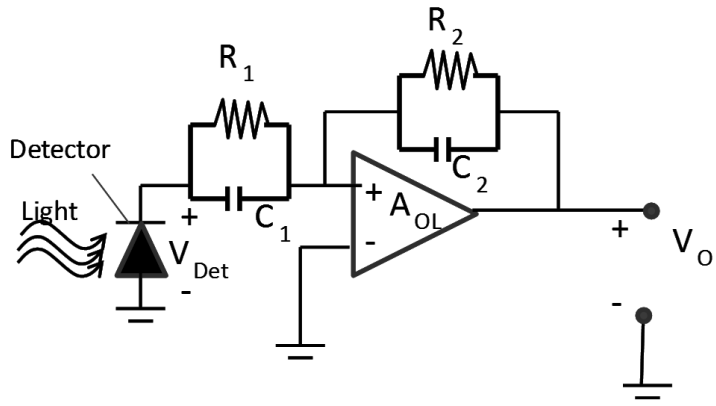


Figure 9.4 Example detector circuit.

where ω is the angular electrical frequency. The voltage gain of the amplifier is

$$A_v = \frac{-Z_2}{Z_1} = -\frac{\dfrac{R_2}{1+j\omega R_2 C_2}}{\dfrac{R_2}{1+j\omega R_1 C_1}} = \frac{-R_2}{R_1}\frac{1+j\omega R_2 C_2}{1+j\omega R_1 C_1}$$

Note that when the frequency is small, the voltage gain is $-R_2/R_1$, and when the frequency is large, the voltage gain is $-C_1/C_2$.

The photoconductive detector provides a change in conductivity with incident photons. The change in conductivity is inversely proportional to the change in resistivity, so the detector can be modeled as a resistor. Therefore, the detector has a variable resistance that changes in an inverse proportional manner to the number of incident photons. The detector circuit is shown in Figure 9.5.

If the circuit is operated in constant source current mode [1] and $R_L >> R_{Det}$, the output voltage is proportional to the change in detector resistance:

$$V_S = \frac{V\Delta R_{Det}}{R_L + R_{Det}} \quad [\text{V}] \tag{9.6}$$

Equation (9.6) can be written in terms of the detector parameters given a sinusoidal irradiance incident on the detector:

$$V_S = \frac{V\Delta R_{Det}}{R_L + R_{Det}} \frac{A_d \eta_q E_{Det}\, \tau}{N(1+\omega^2 t^2)^{1/2}} \quad [\text{V}] \tag{9.7}$$

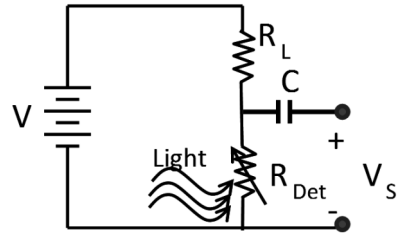


Figure 9.5 Photoconductor circuit.

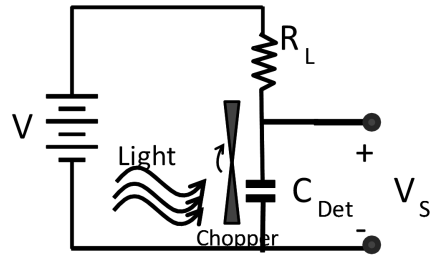


Figure 9.6 Pyrometer circuit.

The detector characteristics are as follows: Ad is the area of the detector, η_q is the quantum efficiency of the detector in converting photons to electrons, E_{Det} is the incident photon flux on the detector in photons per square centimeter per second, t is the average free-carrier lifetime, N is the total number of free charges in the element when there is no incident photon flux, and ω is the angular modulation frequency. The circuit has a capacitor on the output so that the circuit is ac coupled. That is, the output voltage only represents the changes in the detector resistance. The capacitor can be removed to provide a signal proportional to the detector incident flux.

The bolometer can also be modeled as a resistor, but the resistance does not change with a photon-level conversion. The resistance is a measure of resistive material temperature (see the discussion of thermal detectors in Chapter 8). At any rate, the circuit shown in Figure 9.5 can also be used for the bolometer circuit. The capacitor defines the time constant of the system and must be sized accordingly, taking into account the desired cutoff frequency. The pyrometer is modeled with a capacitor, where the capacitance changes with the temperature of the capacitor. To detect the change in capacitance, the light is chopped with rotating blades. A typical circuit for the pyrometer is shown in Figure 9.6.

The photovoltaic circuits shown adequately amplify the input signals, but the other circuits discussed require additional amplification. This can be accomplished by placing the outputs of the circuits at the terminals of the photovoltaic detector. That is, the output of the bias circuits in Figures 9.5 and 9.6 can be considered inputs to the photovoltaic amplifiers in Figures 9.2 and 9.3 (i.e., either the voltage or current mode). This first amplifier is usually called the *preamplifier* because there are other amplifiers in the total system electronics. The preamplifier is the most important circuit in the entire electronics network. The noise associated [2] with the first stage in a network has the predominant effect on the overall noise figure for the network. The system designer should always try to minimize the noise produced in the first stage.

9.2 Conversion of Spatial and Temporal Frequencies

For scanned detectors, there is a relationship [3] between the electronic temporal frequency and the system spatial frequency. This relationship depends on the scan velocity of the detector across the image (or the image velocity across the detector). There are two conversions: One is a function of metric spatial frequency (cycles per

meter) and the other is a function of angular spatial frequency (cycles per milliradian). Consider the cases shown in Figure 9.7. Regardless of whether the detector is serial scanned or parallel scanned, it takes a particular amount of time for the detector to traverse the image. In the first case, the detector traverses the horizontal FOV in a time t_{scan}. The conversion from spatial frequency to electronics temporal frequency is

$$f_e\,[\text{Hz}] = \xi[\text{cycle/mrad}]\frac{FOV\,[\text{mrad}]}{t_{scan}\,[\text{sec}]} \quad [\text{Hz}] \tag{9.8}$$

The relationship is usually inverted to give the electronics transfer function in terms of spatial frequency.

The second case of Figure 9.7 shows a detector scanning across the image plane, where the image plane dimensions are given in meters. Sometimes it is convenient to work in the spatial frequency domain where spatial frequency is given in cycles per meter in the image plane. For this case, the conversion is

$$f_e\,[\text{Hz}] = \xi\,[\text{cycle/m}]\,\frac{\text{Image Size}[\text{m}]}{t_{scan}\,[\text{sec}]} \quad [\text{Hz}] \tag{9.9}$$

In general, the electrical frequency f_e can be expressed in terms of the scan velocity

$$f_e = v\xi \quad [\text{Hz}] \tag{9.10}$$

where the velocity is in meters per second or milliradians per second and the spatial frequency is in cycles per meter or cycles per milliradian, respectively. One of the basic understandings of an EO engineer is the relationship between spatial and temporal frequencies.

The scan velocity can also be written in terms of the detector angular subtense (DAS) and the detector dwell time:

$$v = \frac{DAS}{t_{dwell}} \quad [\text{m/s or mrad/s}] \tag{9.11}$$

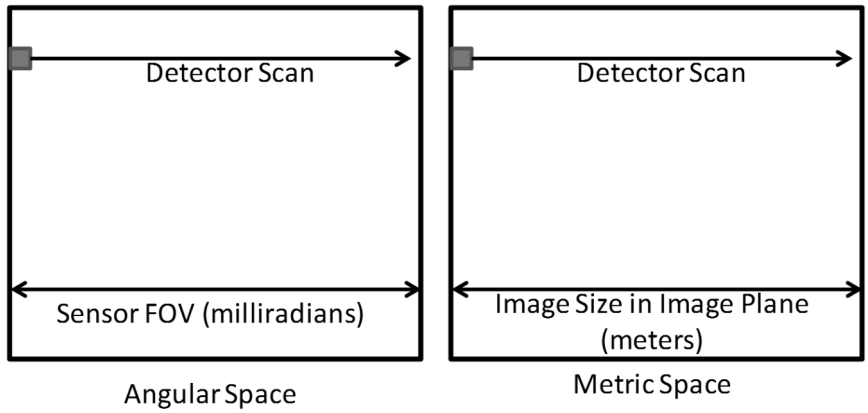


Figure 9.7 Detector scans across image space.

The DAS is the detector size in the scan direction divided by the system focal length. DAS is the angle subtended by the detector onto the angular object space. The dwell time (see Chapter 8) is the time required for a target edge to scan across the detector element.

Example 9.2
The boards and spaces on a picket fence subtend a fundamental spatial frequency of 1.5 cycle/mrad (as seen from a sensor's point of view) in the horizontal scanning direction. The sensor has a 5-deg (87.3-mrad) FOV and operates at 30 frame/s. Assuming a 100% scan efficiency, determine the electrical temporal frequency on the output of the detector that corresponds to the picket-fence fundamental frequency. The solution is given by (9.10) if the scan velocity can be determined. With a 100% scan efficiency, the 30 frame/s requires that (87.3 mrad) × 30/s or 2,618 mrad/s is the scan rate. With a picket-fence fundamental of 1.5 cycles per milliradian, we get

$$f_e = (2.618 \text{ mr/s})(1.5 \text{ cycle/mrad}) = 3,927 \text{ Hz} \tag{9.12}$$

or just under 4 kHz. Higher spatial frequencies require higher electronic frequencies, and lower spatial frequencies require lower electronic frequencies.

Now that temporal frequencies can be converted to spatial frequencies, we can address the image blur associated with the detector integration time as the detector is scanned across the FOV. The detector integration (sometimes called *detector sample-and-hold*) can be performed on the detector chip or off the detector chip on a circuit board. The detector integration circuit sums the detector signal for some integration time and then is reset for the next integration time. Image blur due to this integration is a rectangular function only in the scan direction with a blur size, in time, equal to the integration time t_{int}. The temporal blur (in milliradians) is converted to a spatial blur with the scan velocity, $\alpha_{int} = vt_{int}$. The spatial blur due to detector integration time is present only in scanned imaging systems (not staring systems because the detector remains stationary for the integration time) and can be described with an impulse response:

$$h_{int}(x,y) = \frac{1}{\alpha_{int}} \, rect\left(\frac{x}{\alpha_{int}}\right) \tag{9.13}$$

where the corresponding detector integration transfer function is

$$H_{int}(\xi,\eta) = sin\, c(\alpha_{int}\xi) \tag{9.14}$$

The image blur only occurs in the scan, or horizontal, direction. The function in (9.13) can be considered a separable function where $h(y) = \delta(y)$. The transfer function is also separable where $H(\eta) = 1$.

9.3 Electronics Transfer Function

The electronics transfer function for a particular sensor describes the frequency limitations of the electronic amplifiers and the corresponding filters. Usually, the electronics transfer function provides bandwidth well beyond any other component

in the system, so filters can be added to bound the noise content without reducing system performance. These transfer functions apply primarily to scanned systems since staring systems have an integration time associated with collection of photons on the detector. We continue to provide the scanned electronics functions in this book since because many scanned systems were still in use at the time of this writing.

A common method for modeling the electronic filter transfer functions is to consider the low-frequency response separately from the high-frequency response. The independent transfer function variables are then converted from temporal frequency to spatial frequency and applied as components of the horizontal, or scan direction, transfer function of the system.

A typical low-pass amplifier filter design [4] is the lagging (output voltage lags the input voltage) low-pass single-pole filter like that shown in Figure 9.8(a). The transfer function for this filter can be written in terms of its single-pole response:

$$H(f_e) = \frac{P}{f_e + P} \quad \text{[unitless]} \tag{9.15}$$

where the pole is $p = 1/(2\pi RC)$. With some mathematical manipulation, the magnitude of the transfer function can be written as

$$H(f_e) = \frac{1}{\sqrt{1 + \left(\frac{f_e}{P}\right)^2}} \quad \text{[unitless]} \tag{9.16}$$

The magnitude of the transfer function for this filter is shown in Figure 9.8(b). The magnitude in decibels can be calculated by

$$\left|H(f_e)\right|_{dB} = 20\ log\left|H(f_e)\right| \quad \text{[unitless]} \tag{9.17}$$

The transfer function can be easily converted to the spatial frequency domain for application as a system component simply by modifying p to a spatial pole $1/(2\pi RC)$ and f_e with ξ in (9.15). Also note that v is the scan velocity in milliradians per second. The MTF for the filter can then be written as

$$H_{LowPass}(f_e) = \frac{1}{\sqrt{1 + \left(\frac{\xi}{p}\right)^2}} \quad \text{[unitless]} \tag{9.18}$$

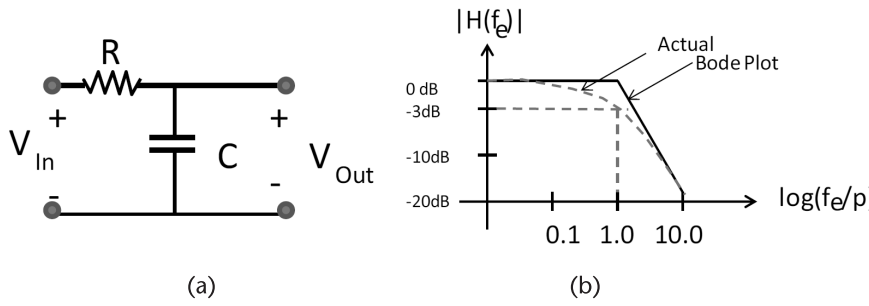


Figure 9.8 Low-pass filter: (a) circuit and (b) transfer function.

The order of the filter can be increased using the same filter cascaded (i.e., the output of one filter feeds the input of the next filter). These are called *ladder filters*. A large number of different ladder circuits have been developed, including both Butterworth and Chebyshev filters. The spatial frequency equivalent MTF corresponding to the magnitude function [5] for the *n*th-order Butterworth filter is

$$H_{LowPass}(f_e) = \frac{1}{\sqrt{1+\left(\frac{\xi}{p}\right)^{2n}}} \quad \text{[unitless]} \tag{9.19}$$

A similar filter can be designed for high-pass applications. This is the case for ac-coupled systems were the dc signals are removed from the output of the detector. In some cases, dc is added back to the ac signals for later display. These types of systems are called *dc-restored systems*. A simple, single-pole, high-pass filter is shown in Figure 9.9(a) and the corresponding Bode plot is shown in Figure 9.9(b). The circuit shown in Figure 9.9 is a leading high-pass filter where the transfer function can be written in terms of the pole:

$$H(f_e) = \frac{f_e}{f_e + p} \quad \text{[unitless]} \tag{9.20}$$

where the pole is again at $p = 1/(2\pi RC)$. We are more interested, however, in the magnitude of the signal as it transfers through the system.
With some algebra, the magnitude of the function given in (9.20) is

$$H(f_e) = \frac{\frac{f_e}{p}}{\sqrt{1+\left(\frac{f_e}{p}\right)^2}} = \frac{1}{\sqrt{1+\left(\frac{p}{f_e}\right)^2}} \quad \text{[unitless]} \tag{9.21}$$

There are a large number of high-pass filters, with the Butterworth and Chebyshev filters being the more common ones. The transfer functions can be converted to the spatial frequency domain by changing the temporal frequency to a spatial

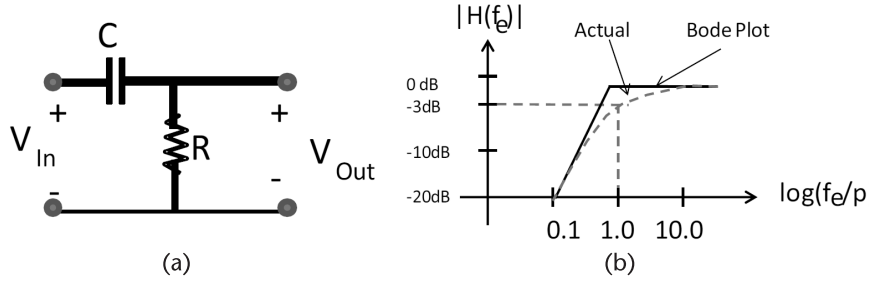


Figure 9.9 High-pass filter.

frequency along with changing the pole to a spatial frequency with the scan velocity term $p = 1/(2\pi RC)$. The high-pass filter MTF for an nth-order Butterworth filter is

$$H_{LowPass}(\xi) = \frac{1}{\sqrt{1+\left(\frac{p}{\xi}\right)^{2n}}} \qquad \text{[unitless]} \tag{9.22}$$

The combination of the high pass and the low pass can be implemented simply by multiplying their MTFs. In circuit design, the output of one filter feeds the input of the other filter. From Chapter 3, we know that in the time domain, the input signal of the first filter is convolved with the filter impulse response to result in the filter output signal. The output signal is then convolved with the impulse response of the second filter. The output of the second filter is the output of the filter system. Care must be taken in the characterization of the loading effects at each stage of analysis. In most cases, it is easier to work in the frequency domain where the transfer functions multiply.

9.4 Noise

Entire books have been written on noise analysis in circuits. A complete treatment of circuit noise here is beyond the scope of this work; however, a cursory overview of the noise associated with detector electronics is appropriate. The noise components associated with the detector are described in Chapter 8. Noise components that are specific to EO and I^2R systems are presented later in this book. The noise components associated with the detector circuit and system electronics are Johnson, $1/f$ (or flicker noise), shot, microphonics, readout, and quantization noise. The primary noise components of Johnson, I/F, and shot noise are addressed here.

Johnson Noise

Johnson noise [6] exists because all conductors exhibit thermal noise. Any material over the temperature of 0 K exhibits electrical noise from the thermal activity of the electron states. A resistor at some temperature T provides a noise level that is related to its resistance and temperature. In theory, capacitors and inductors do not exhibit thermal noise. In practice, even high-quality capacitors and inductors generate some noise. Thermal noise sources associated with resistors can be modeled as either a voltage source or a current source, as shown in Figure 9.10.

The value of the rms noise voltage for the voltage model shown in Figure 9.10 is

$$v_{Johnson} = \sqrt{4kTR\Delta f} \quad \text{[V]} \tag{9.23}$$

where k is Boltzmann's constant of 1.38×10^{-23} J/K, T is the temperature of the resistor in kelvin, R is the resistance of the resistor, and Δf is the bandwidth over which the noise is measured. The equivalent current value for the second model is

$$i_{Johnson} = \sqrt{4kT\Delta f(1/R)} \quad \text{[A]} \tag{9.24}$$

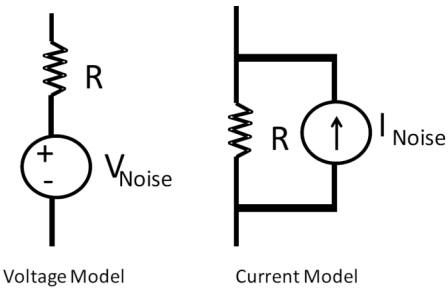


Figure 9.10 Johnson noise models.

Example 9.3
Determine the noise voltage for a 10-kΩ resistor at room temperature (300K). The noise bandwidth is 10 kHz. The solution is given by (9.23), where

$$v_{Johnson} = \sqrt{4(1.38\times10^{-23}\ \text{J/K})(300\text{K})(10{,}000\Omega)10{,}000\text{Hz}} = 1.29\mu\text{V}$$

Note that the rms thermal noise of the resistor is around a microvolt. A reduction in resistance, temperature, or bandwidth decreases the noise voltage.

1/*f* Noise

1/*f* noise, or *flicker noise*, accompanies any system where electron conduction is present [7]. This noise is due to surface imperfections caused during the fabrication process. The noise has a power spectrum that decreases with frequency, hence its name, and is more important at lower frequencies (from 0 to 100 Hz). The long-term drift of transistor amplifiers is usually the result of 1/*f* noise. Flicker noise can be described by an rms current:

$$i_{1/f} = \sqrt{KI_{DC}\frac{\Delta f}{f}} \quad [\text{A}] \tag{9.25}$$

where K is a constant that is specified for a semiconductor device, I_{DC} is the average (i.e., dc) current through the device, and Δf is the bandwidth.

Shot Noise

Shot noise in the electronics is the result of random current fluctuations caused by the discreteness of charge carriers when diffused across a semiconductor junction. The rms value for shot noise current is

$$i_{shot} = 2eI_{DC}\Delta f \quad [\text{A}] \tag{9.26}$$

where e is the charge of an electron, 1.6×10^{-19} C. Both shot noise and Johnson noise have a power spectrum that is approximately constant across the frequency band. These are examples of *white noise*.

There is a difference in the way noise sources add versus typical voltage and current sources. DC voltages and currents are additive, and ac voltages and currents

add vectorally (with amplitude and phase). However, noise sources add constructively. For two-noise voltage sources in series:

$$v_{total}^2 = v_1^2 + v_2^2 + 2\alpha v_1 v_2 \quad [\text{V}] \tag{9.27}$$

where α is the correlation coefficient for the noise sources. For uncorrelated noise sources, which includes the Johnson noise and the majority of other noise cases, the third term is zero and the noise sources add in quadrature. Two identical waveforms have a correlation coefficient of 1. In a few circumstances, noise sources are correlated, as in the case of ground-loop pickups of stray signals. The majority of sensor noise contributors are assumed to be uncorrelated.

There are many other types of noise. Diode noise, BJT noise, and FET noise are all combinations of shot and 1/*f* noise. Other noise sources include power-source leakage, ground-loop pickup, and cable pickup (of electromagnetic fields). Some general guidelines in the noise reduction process are as follows:

- Use filtering to limit the noise bandwidth.
- Match the input resistance of the amplifier to the detector.
- Reduce lead lengths as much as possible.
- Keep the power source clean. Decouple and regulate the power input to each board.
- Minimize any ground impedances. Ground planes are essential to isolate digital circuits from sensitive analog circuits.
- Keep digital grounds separate from analog grounds. They should connect at only one point (the point of lowest potential).
- Isolate high-impedance, high-gain inputs from strong drive currents. This usually requires shielding.
- Minimize magnetic loops. Reduce the loop area of wires, preferably with twisted pairs.

While the minimization of noise is, in general, frustrating, these general guidelines can simplify the noise reduction process.

The *noise bandwidth* is defined differently than a network transfer function. It is not identical to the system bandwidth but describes the characteristics of the noise. The noise bandwidth is the frequency span of a rectangular-shaped power-gain curve equal in area to the area of the actual power gain versus frequency function. Consider the power spectrum [or power spectral density (PSD), $G(f)$, of the noise shown in Figure 9.11.

The noise bandwidth can be calculated by

$$\Delta f_{noise} = \frac{1}{G_o}\int_0^\infty G(f)\,df \quad [\text{Hz}] \tag{9.28}$$

Once the noise bandwidth is calculated, the amount of rms noise voltage in a 1-Hz bandwidth can be estimated:

$$v_{rms} = \sqrt{G_o \Delta f_{noise}} \quad [\text{V}/\sqrt{\text{Hz}}] \tag{9.29}$$

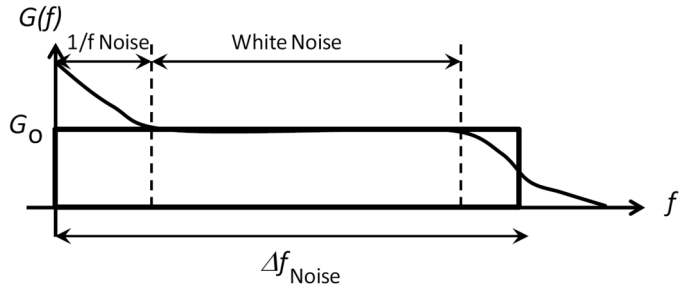


Figure 9.11 Noise power spectrum and noise bandwidth.

Two parameters that describe the noise associated with an amplifier or filter stage are the noise factor and the noise figure. The noise factor is defined as the input SNR normalized by the output SNR:

$$F = \frac{S_i/N_i}{S_o/N_o} \quad \text{[unitless]} \tag{9.30}$$

The noise factor provides the noise figure in decibels. The definition for noise figure is

$$NF = 10 \log F \quad \text{[dB]} \tag{9.31}$$

An amplifier or filter stage with no noise will have an *NF* of zero. A well-designed amplifier should have an *NF* of less than or equal to 3 dB.

The noise factor allows the determination of critical design guidance in the staging of amplifiers and filters. Consider the cascaded network shown in Figure 9.12. Let *G* describe the voltage gain and *F* describe the noise factor for a particular circuit in the cascaded network. The noise factor for the entire cascaded network can be estimated to be

$$F_{123} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} \tag{9.32}$$

The impact of the first stage on overall system noise is higher given that the cascaded gains are larger than 1. Each stage should have a gain greater than 1 with the highest gain occurring at the first stage, where the sensor noise dominates.

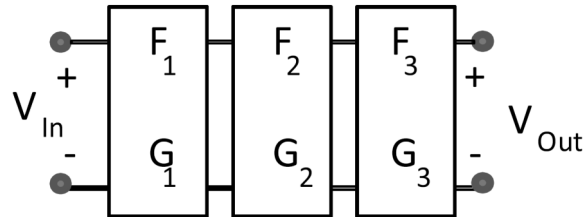


Figure 9.12 Cascaded network.

9.5 MTF Boost Filter

The *MTF boost filter* is an active filter that provides gain in the compensation of MTF rolloff. The boost filter provides a boost to particular spatial frequencies while maintaining a unity transfer for low and high frequencies. Boost [8] is available from a variety of circuits. Here, we present a common representation:

$$MTF_{boost}(f_e) = 1 + \left(\frac{K-1}{2^N}\right)\left[1 - cos\left(\pi \frac{f}{f_{boost}}\right)\right]^N \tag{9.33}$$

where K is the boost amplitude, N is the boost order, f is the temporal frequency of the circuit, and f_{boost} is the frequency that benefits the most from the boost. Figure 9.13 shows the boost gain as a function of temporal frequency for various boost amplitudes with $N = 1$. The boost filter can be converted to spatial frequency using the temporal-to-spatial conversion given in Section 9.2 to yield

$$MTF_{boost}(\xi) = 1 + \left(\frac{K-1}{2^N}\right)\left[1 - cos\left(\pi \frac{\xi}{\xi_{boost}}\right)\right]^N \tag{9.34}$$

The boost filter provides little gain at low frequencies and little gain at high frequencies. At the boost frequency, the signal is boosted by a factor of K. The spatial frequency version has the same shape and boost amplitude with a change in the independent variable.

9.6 EO Mux MTF

The electro-optical multiplexer (EO mux) is an image converter. It provides infrared detector output signals from a linear array and their corresponding amplifiers to

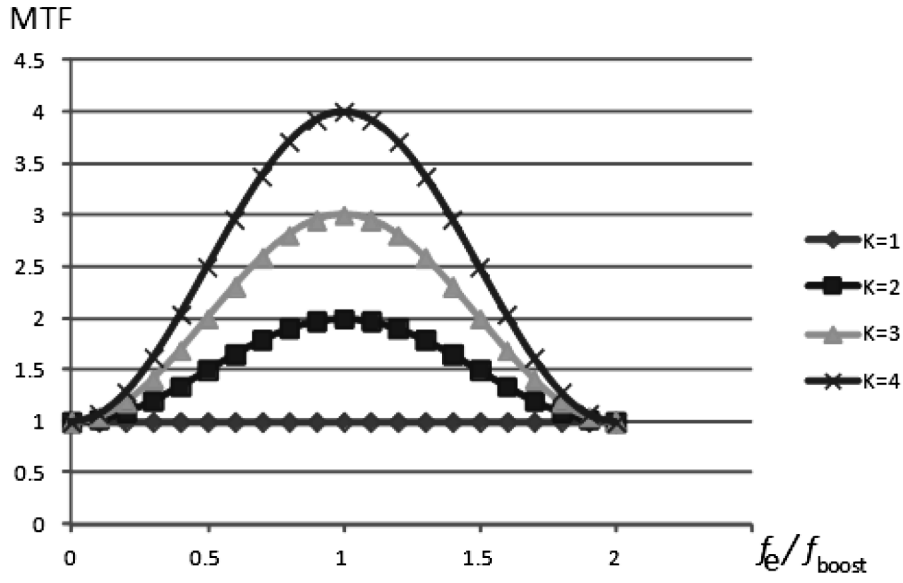


Figure 9.13 MTF boosting filter.

a visible linear light-emitting diode (LED) array. The infrared signals light up the LEDs to an intensity (in the visible wavelengths, hence, EO mux) proportional to the signal voltage. Consider the EO mux shown in Figure 9.14. The EO mux shown has a linear array of five infrared detectors, corresponding electronics, and LEDs to show the concept. In reality, common module FLIRs have 60, 120, or 180 detectors in a linear array.

A 2D image is formed by the infrared lens onto an image plane. The front side of the scanner is reflective in the infrared wavelengths, and the scanner sweeps the image across the linear detector array in one direction. When the edge of the image is reached, the scanner reverses direction and sweeps the image back across the linear array. This backscan is accompanied by a small tilt of the scan mirror in the vertical direction to fill in between the forward scan lines to compensate for the gap between detector elements. The output voltages are taken from the infrared detectors, amplified, and fed to the visible LED array. The LED array is imaged by a visible lens and reflected off the backside of the scanner converging to a visible image plane. The image plane can be viewed directly by a human, or it can be imaged onto a vidicon tube for conversion to video output.

The LED spatial extent has an MTF associated with it. The EO mux display is continuous in the horizontal direction with a continuous voltage out of each diode. In the vertical direction, the output is sampled as there are a discrete number of LEDs, just as the detectors sampled the original infrared image. The MTF treatment of the EO mux LED spatial characteristics is discussed in Chapter 10.

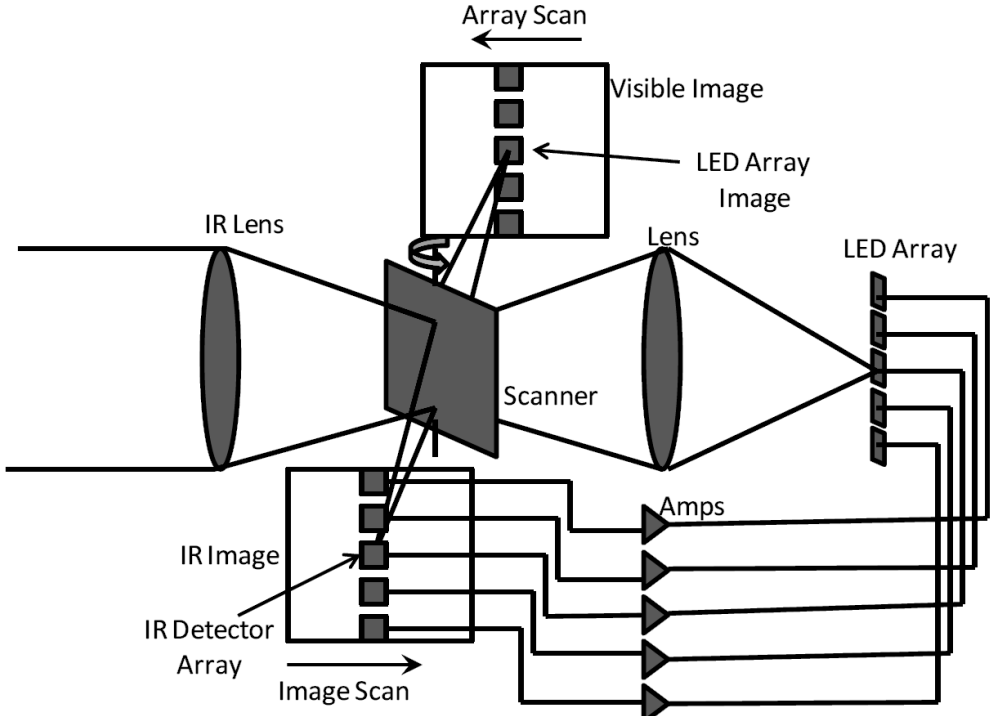


Figure 9.14 EO multiplexer.

9.7 Digital Filter MTF

The two major classes of digital filters are *infinite impulse response* (IIR) and *finite impulse response* (FIR) filters. IIR filters have excellent amplitude responses but are more difficult than FIR filters to implement. FIR filters usually have more ripple than IIR filters. We focus here on FIR filters. We do not derive the transfer functions for these digital filters because these derivations are covered elsewhere (e.g., [7]). However, we are interested in the transfer functions as they affect the overall performance of our system.

The filters are shown in Figure 9.15 operating on a linear stream of pixel values. The output pixel signal level of the odd filter, in the case shown, depends on the value of five input pixels. The input pixel signal levels are multiplied by their corresponding coefficients and the products are summed to give an output pixel signal level. The input pixels are shifted by one pixel and the output pixel signal level is again calculated. This process continues through the stream of data. There are artifacts at data stream edges (e.g., frame- or line-scan edges), where tricks such as doubling back data have to be played to reduce the artifacts.

Note that the sum of the filter coefficients must be 1. Otherwise the average frame gain and offset change. The filter MTF in the filtering direction for an odd number of samples [9, 10] is

$$MTF_{dfilter}(\xi) = \sum_{k=1}^{N/2} A_K \cos\left[\frac{2\pi k\,\xi}{\xi_s}\right] \tag{9.35}$$

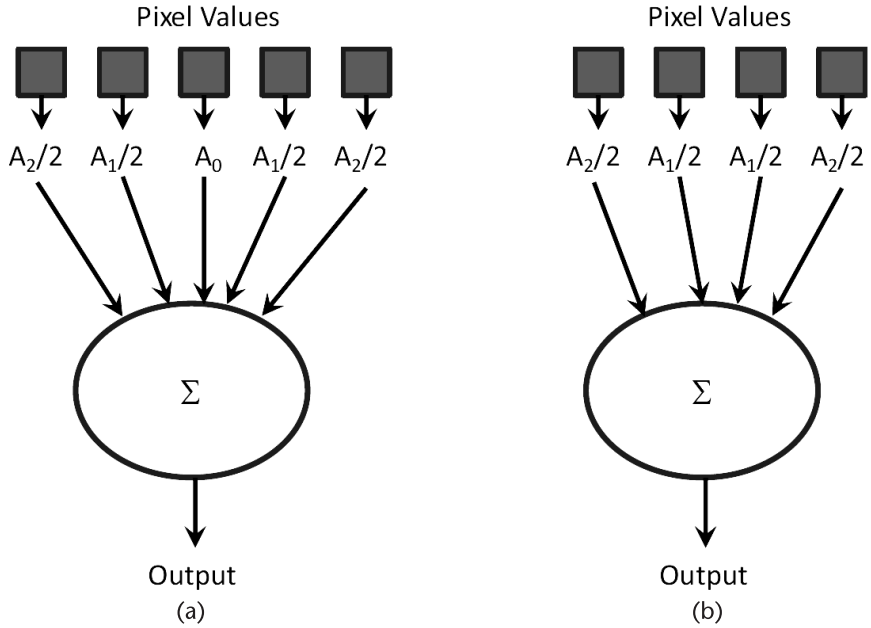


Figure 9.15 Filters with (a) odd and (b) even numbers of samples.

where $\xi_S = 1/S$ is in cycles per milliradian and S is the angular distance between pixel samples in milliradians. Note that with oversampling, S may be smaller than a DAS. The MTF for an even number of samples filter is

$$MTF_{dfilter}(\xi) = \sum_{k=1}^{N/2} A_K \cos\left[\frac{\pi(2k-1)\,\xi}{\xi_s}\right] \tag{9.36}$$

While these filters are written for one direction, they can be easily applied in the horizontal or vertical direction if separable (see Chapter 3) functions are used. Nonseparable 2D filters are much more difficult to analyze.

9.8 CCDs

Some visible and infrared staring arrays are CCDs. These CCDs comprise an array of detectors that sum the irradiance on the detectors for some integration time. The integration time is less than the frame time of typically around 1/30th of a second. The integrated charge is then moved out of the CCD along a predetermined path [11] under the control of clock pulses. The charge is stored and moved in a series of MOS capacitors. Adapted from *The Infrared Handbook* [12], Figure 9.16 shows a typical three-phase readout cycle.

The electrode potentials at four different times are shown to depict the charge transfer. At time t_1, a low potential is held across the V_1 electrodes. At this time, the charge packets are located at the V_1 positions. A low potential is placed on the V_2 positions at t_2 and the potential is increased at the V_1 positions. This process is increased at t_3 causing even more of the charge to transfer to the V_2 positions. Finally at t_4, the only low potentials are located at the V_2 positions.

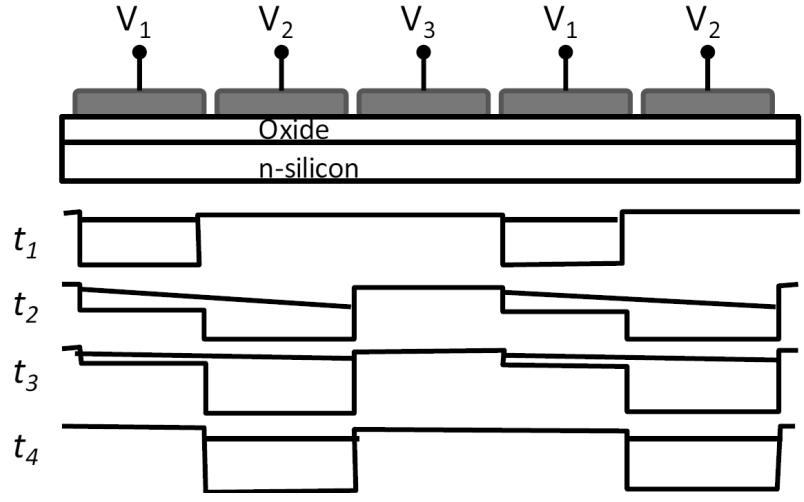


Figure 9.16 Three-phase CCD operation.

This same process is followed to cause the charge packets to transfer to the V_3 positions.

The process described above has a transfer efficiency that describes the completeness of charge packet transfer. Charge is typically left behind, causing a smearing effect. Holst [7] describes an MTF for the smearing effect:

$$MTF_{transfer} = exp\left\{-N_{trans}(1-\varepsilon)\left[1-cos\left(\frac{2\pi\,\xi}{\xi_{ts}}\right)\right]\right\} \tag{9.37}$$

N_{trans} is the total number of charge transfers from a detector to the output amplifier, ε is the transfer efficiency, and ξ_{ts} is the clocking spatial frequency. For horizontally clocked packets, where the horizontal FOV is much larger than a DAS,

$$\xi_{ts} \approx \frac{N_{trans}}{HFOV} \tag{9.38}$$

Holst also points out that an average is half the maximal number of transfers. Transfer efficiencies range from 0.99 to 0.99999.

9.9 Uniformity Correction or "NUC"

In systems that have more than one detector, such as a scanned linear array or a staring array, more than one detector output signal provides the image. In these systems, the combination of the detectors, filters, and amplifiers may provide images that are not uniform in gain or offset. That is, the output images do not provide signals that are *uniform* functions of input irradiance. The output image in terms of gain and uniformity can be written as a function of the input image on the detector array:

$$O(x,y) = I(x,y)\,M(x,y) + B(x,y) \tag{9.39}$$

where $M(x,y)$ is the gain and $B(x,y)$ is the offset. Equation (9.39) provides the equation for a line that gives the gain and offset for a particular pixel. Sometimes a *gamma* correction is required to force the throughput to be a line instead of a nonlinear input intensity-dependent gain and offset function. The gain can be considered the amplification of the signal and is sometimes known as *contrast* (in television systems). The offset is a standard background signal in order to control the level (or *brightness*). The gain depends on a number of contributors, including detector responsivity, amplifier gain, and filter calibration. The offset is more of an electronic source potential for ac-coupled systems and a combination of source potential and other contributors for dc-coupled systems.

For systems with staring arrays, each detector has a unique responsivity and some unique electronics. Staring arrays have been noted to be very nonuniform, especially in the infrared wavelengths. A number of techniques are currently under investigation to correct for staring array nonuniformities, including cali-

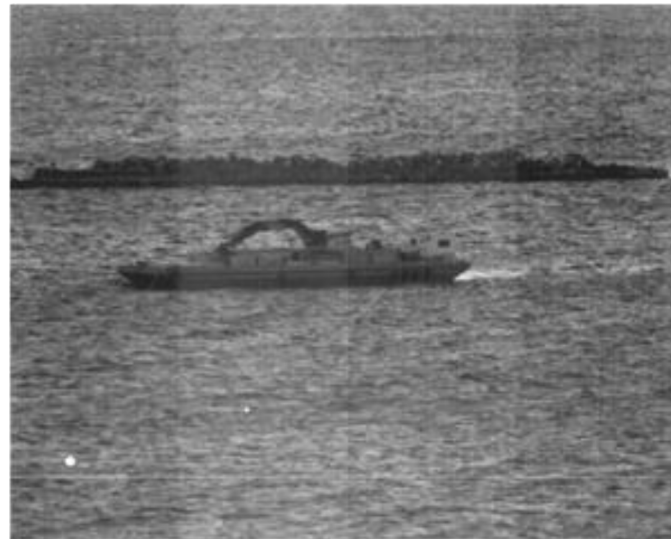


Figure 9.17 Nonuniform SWIR image.

bration and correction electronics. More extravagant solutions include retinal processors that adaptively correct the gain and offset based on the input imagery [13].

For scanned linear arrays, the detectors and electronics are common for each row. Therefore nonuniformities are seen more in line-to-line discrepancies. The correction for scanned imagers is typically a bank of amplifier and offset voltage adjustments or digital gain and offset equivalents.



Figure 9.18 Corrected SWIR image.

As an example, Figure 9.17 is a SWIR image that was collected before the nonuniformity correction (NUC) was applied. In this case one sees vertical striping associated with the variability in camera electronics. The image in Figure 9.18 shows that this effect has been completely eliminated from the scene. Although more sophisticated NUC schemes could be implemented (e.g., three-point polynomial fit) most nonuniformity can be sufficiently removed with a linear, two-point correction.

9.10 Readout Integrated Circuits

One of the biggest design challenges in camera system electronics is how to transfer the signal (image) generated at the focal plane array off the camera (e.g., to a monitor or storage device). Accomplishing this task requires an appropriately designed readout integrated circuit. The job of the modern-day ROIC is to take the charge generated at the detector, condition the signal (e.g., charge-voltage conversion, amplify, antiblooming), apply any additional processing (e.g., band limiting, buffering), multiplex the signals coming off each individual unit cell, and send the signal downstream for digitization and display. Moreover, much of this functionality needs to be accomplished in a very small area, because modern-day unit cells are on the order of tens of microns a side. A schematic of a modern-day ROIC is shown in Figure 9.19.

As indicated in Figure 9.19, the detector array must be physically coupled to the ROIC, a process referred to as *hybridization*. Hybridization must be accomplished in such a way as to provide good electrical contact between ROIC and FPA, yet minimize the mechanical/thermal stresses that result when two dissimilar materials are bonded. This latter issue, if not accounted for, can result in the array delaminating from the ROIC, thus significantly degrading camera

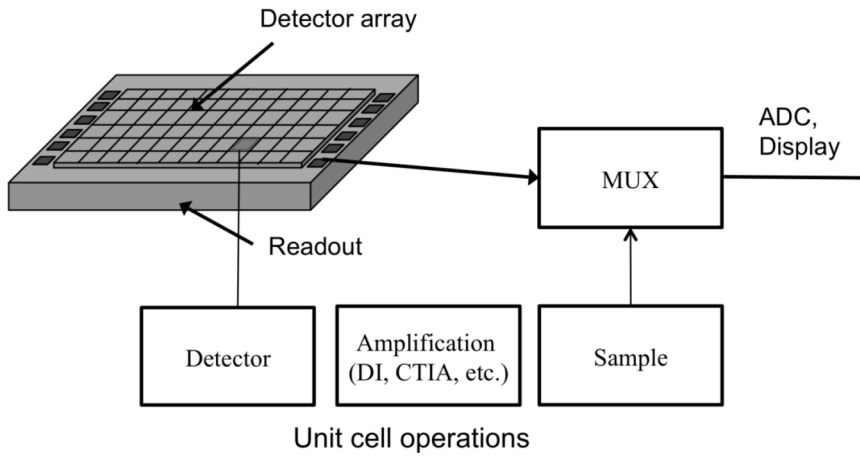


Figure 9.19 Schematic of a modern-day ROIC. Each unit cell detects, amplifies, and samples the incident radiation after it has been converted to a voltage. The ROIC multiplexes the signals from all unit cells and passes this signal off-chip for digitization and display. Some newer ROICs even perform the digitization on-chip.

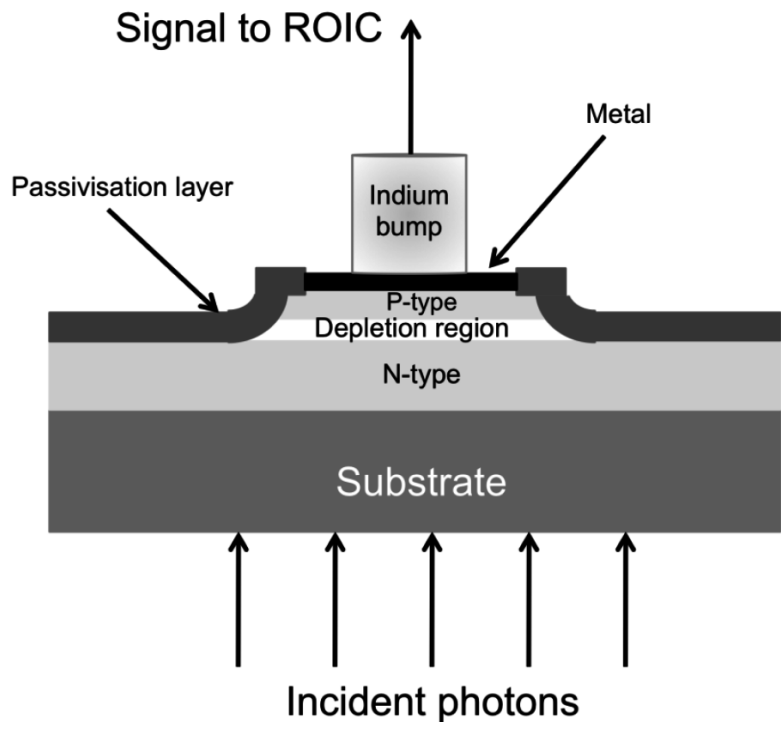


Figure 9.20 Single photovoltaic detector. Incident photons pass through the detector substrate and generate a charge at the p-n junction. The signal is physically coupled to the ROIC via an indium bump. Bonding the ROIC to the FPA requires very high pressures, good "bump" quality (ideally flat on top), and good alignment between detector array and ROIC.

performance. Probably the most popular approach to fabrication is the *bump bonding* approach whereby each unit cell in the detector array is connected to the ROIC through a small amount of indium, chosen for its good mechanical and electrical properties. A typical photovoltaic detector sensing element is in shown in Figure 9.20.

Incident photons pass through the detector substrate (e.g., CdZnTe, Si) and are absorbed by the "N-type" material whereby a current is formed across the p-n junction. A passivation layer is typically applied to the detector surface to minimize carrier charge recombination at the detector surface and this layer provides good electrical stability. The signal-generated current then flows through the indium contacts to the ROIC for further processing. Successfully bonding the detector element to the ROIC requires extremely high pressures (several thousand psi) and/or temperatures [14]. The process also requires submicron tolerances in aligning the indium bumps on the detector array with the corresponding bumps on the ROIC. Despite these difficulties, sensor manufacturers can now achieve >99% pixel operability even with small (<20 μm) pixel sizes and large-format (1,024 × 1,024 pixel) arrays.

Once good electrical and mechanical coupling has been achieved, we can proceed to define the specifics of the "unit cell operations"also shown in Figure 9.19. The specifics of a particular ROIC design will depend, as always, on the system requirements.

In all cases we define a "good" ROIC design to be one that maximizes the SNR for the desired bandwidth and dynamic range of the sensor. A ROIC must contend with a number of noise sources including white (thermal) noise, dark current, errors in resetting capacitors used in sampling, and $1/f$ noise [15]. Different sensor manufacturers deal with these errors in different (often proprietary) ways. Historically, the resistor transimpedance amplifier (RTIA) circuit was used to condition/amplify signals coming off the detector. However, this particular circuit requires more space than is available in many of today's small-pitch FPAs. Moreover, the high resistance required of this circuit can result in unacceptably high levels of $1/f$ noise for low-flux applications. More common approaches to signal amplification for today's systems are direct injection (DI) amplification and capacitive transimpedance amplification (CTIA) [15]. Both work well in low-flux (low-signal) environments; however, the latter is the more commonly used due to its good sensitivity and dynamic range performance.

With regard to sampling strategies there are two primary approaches: sample-and-hold (SH) and correlated double sampling (CDS) [16]. The former performs as the name suggests: A capacitor integrates charge until a gate is closed, thereby generating the sampled output. The integrator acts as a low-pass filter and thus has the benefit of rejecting (averaging) some of the high-frequency noise sources. However, low-frequency noise remains that is associated with resetting (discharging) the capacitor. The idea behind CDS is to sample the signal twice during a given time interval and use the difference as the output. This has the desired effect of reducing the low-frequency readout noise; however, this is achieved at the cost of amplifying higher frequency noise. Nonetheless, CDS is a commonly used approach in many of today's staring array IR cameras.

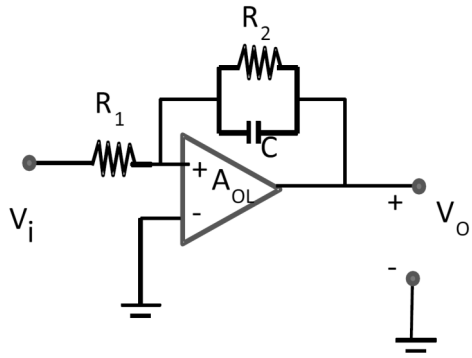
Of course, there are limits to how long a signal can be integrated prior to saturating. One of the jobs of the ROIC is to prevent pixel saturation for large-flux signals, thereby increasing the dynamic range of the camera. The saturation limit is referred to as the *well capacity* of the pixel and is largely governed by the capacitance of the integrating circuit. Circuits with larger capacitance hold more charge, hence more signal; however, large capacitors require more physical space, which can be a serious constraint for modern FPAs. Moreover, for low light levels a large well capacity can be a hindrance because the entire signal may be compressed into a small portion of the full dynamic range, resulting in poor sensitivity. Different strategies exist for how to maintain good performance in both low- and high-flux environments. One possibility is to have multiple capacitors in a given circuit that can be instantaneously switched depending on the incoming signal; that is, low light levels use a smaller capacitor than high light levels. Another is to automatically adjust the integration time to match the incoming signal, with large signals being assigned shorter integration times. Modern-day ROICs have well capacities well into the hundreds of thousands of electrons and can maintain good sensitivity and noise performance across this range.

Designing and manufacturing high-quality ROICs is an ongoing challenge for camera manufacturers. The need for increasingly large format arrays with smaller pixels requires nontrivial adjustments to both the design of the individual unit cells

and the tools required for assembly. For this reason, large-format, small-pixel arrays remain an active area of research and development.

9.11 Exercises

9.1 Determine the transfer function and the 3-dB cutoff frequency for the detector amplifier shown below. Assume an ideal operational amplifier.



9.2 A HgCdTe photovoltaic detector has a resistance of 10 kΩ. Design a voltage-mode detector amplifier for the detector with a bandwidth of 5 Hz to 1 MHz and a closed-loop, midband gain of 30 dB.

9.3 A common-module FLIR system has 180 detectors that scan in the horizontal direction at a rate of 60 times per second (30 Hz and a 2:1 interlace pattern). The scan efficiency is 0.8 (meaning that 20% of the time the scanner is performing a task other than scanning the image across the detector turnaround time, stop time, retrace time, and so forth) If the horizontal FOV is 2.75 deg and the detector scans across a picket fence with equally spaced boards 1 mrad apart, what is the (a) scan velocity and (b) temporal signal frequency corresponding to the fundamental harmonic of the fence?

9.4 Consider two resistors (R_1 = 50 kΩ and R_2 = 20 kΩ) that are connected in series. At a temperature of 400 K and a bandwidth of 5 MHz, calculate the rms noise voltage across each resistor and then across both resistors.

9.5 A system input SNR ratio is 100 and the system output SNR is 30. Determine the system noise factor and noise figure.

9.6 Plot the MTF for a boost filter with a K of 4 and a boost temporal frequency of 2 MHz. Assume a 50-mrad HFOV and a 52-μs HFOV scan time.

9.7 An odd digital filter is described with the following coefficients: $A_0 = 0.8$, $A_1 = 0.9$, $A_2 = -0.2$, and $A_3 = -0.5$. Sketch the MTF in terms of the normalized spatial frequency ξ/ξ_0.

9.8 Plot the MTF as a function of normalized spatial frequency (normalized to the clocking spatial frequency) for CCD transfer efficiencies of 0.99, 0.999, 0.9999, and 0.99999. Assume the number of transfers to be 300.

References

[1] Wolfe, W., and G. Zissis, *The Infrared Handbook*, Environmental Research Institute of Michigan, Office of Naval Research, Washington, DC, pp. 11–32, 1993.

[2] Krauss, H., C. Bostian, and F. Raab, *Solid-State Radio Electronics*, New York City, NY: Wiley, 1980.

[3] Holst, G., *Electro-Optical Imaging System Performance*, Orlando, FL: JCD Publishing, 1995, p. 119.

[4] Valkenburg, M., *Analog Filter Design*, New York City, NY: Holt, Rinehart, and Winston, 1982, p. 80.

[5] Sedra, A., and K. Smith, *Microelectronic Circuits*, New York City, NY: Holt, Rinehart and Winston, 1991, p. 771.

[6] Watts, R., *Infrared Technology Fundamentals and System Applications: Short Course*, Environmental Research Institute of Michigan, 1990.

[7] Holst, G., *Electro-Optical Imaging System Performance*, Orlando, FL: JCD Publishing, 1995, p. 146.

[8] Savant, C., M. Roden, and G. Carpenter, *Electronic Circuit Design*, Menlo Park, CA: Benjamin Cummings Publishing, 1987, p. A125.

[9] FLIR92 Thermal Imaging Systems Performance Model, U.S. Army Night Vision and Electronic Sensors Directorate Report, Jan. 1993.

[10] Rabiner, R. L., and B. Gold, *Theory and Application of Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975, pp. 81–84.

[11] Streetman, B. G., *Solid-State Electronic Devices*, Englewood Cliffs, NJ: Prentice-Hall, NJ, 1980.

[12] Wolfe, W., and G. Zissis, *The Infrared Handbook*, Environmental Research Institute of Michigan and the Office of Naval Research, 1993.

[13] Scribner, D. A., K. A. Sarkady, M. R. Kruer, and J. T. Caulfield, "Adaptive Retina-Like Preprocessing for Imaging Detector Arrays," *IEEE International Conference on Neural Networks*, Washington DC, 1993.

[14] Efimov, V. M., and D. G. Esaev, "Method for Determining Critical Pressure in Assembling Hybrid MCT Focal Plane Arrays," *Optoelectronics, Instrumentation and Data Processing*, Vol. 43(4), 2007, pp. 370–374.

[15] Hoffman, A., and J. Vampola, "Modern Infrared Detectors and Systems Applications" short course notes, University of California, Santa Barbara, 2011.

[16] Johnson, J. F., and T. S. Lomheim, "Focal-Plane Signal and Noise Model-CTIA ROIC," *IEEE Transactions on Electron Devices*, 56(11), pp. 2506–2515, 2009.

CHAPTER 10

Image Processing

In this chapter we review some of the basics of image processing and provide a brief overview of some of the more advanced techniques. Although it is not possible to provide a comprehensive look at the field in a single chapter, our aim is to provide a practical guide to image processing as it relates to EO/IR system design. The focus will be on the influence of sampling on digital image fidelity and on methods for overcoming sampling limitations. For each of the topics covered we will refer the reader to the appropriate reference for further study.

The basics of sampling theory are briefly reviewed in Section 10.1 along with some basic filtering approaches for eliminating sampling artifacts. Section 10.2 discusses some common filtering applications: image interpolation, localized contrast enhancement, and "boost." In Section 10.3 we review super-resolution techniques that can overcome some of the limitations of sampling theory. Finally, in Section 10.4 we provide an overview of "image fusion" techniques whereby data (images) from multiple sensing modalities are combined to provide the end user with a synthesized image that contains information from each of the individual modalities. We summarize in Section 10.5.

10.1 Basics of Sampling Theory

The goal in sampling is to make discrete observations of a continuous (analog) signal in a manner that permits accurate reconstruction of that signal. For example, an imager's focal plane array spatially discretizes electromagnetic field intensity to produce an image, where the spacing between samples in the x and y spatial dimensions is given by Δ_x and Δ_y μm, respectively. The ability of the practitioner to accurately recover the true underlying image from these finite samples depends on (1) the image content and (2) the interpixel spacing. Claude Shannon provided specific conditions for recovery in the late 1940s. Shannon's paper, "Communication in the Presence of Noise," showed that if a continuous function contains temporal frequencies no greater than $1/2\Delta$ cycle/s, then that function may be exactly reconstructed by a series of observations spaced Δ sec apart [1]. Moreover Shannon specified how to perform the reconstruction. In general, given a one-dimensional function $g(x)$ sampled at discrete points $g(n\Delta_x)$ $n = \ldots, -1, 0, 1, \ldots$, where Δ_x is the intersample spacing, the function $g(x)$ is given by

$$g(x) = \sum_{n=-\infty}^{\infty} g(n\Delta_x)\, sinc\left(\frac{x - n\Delta x}{\Delta x}\right) \tag{10.1}$$

For example, assume that a one-dimensional image $g(x)$ contains spatial frequencies no greater than $1/2\Delta_x$ cycle/μm or cycle/mrad (the interpixel

spacing is often quantified by the angle subtended by an individual pixel, hence units are often specified as mrad^{-1}). If the interpixel spacing is no greater than Δ_x, then (10.1) says that the image can be exactly recovered at any point in space using a *sinc*() function to interpolate. The frequency $f_{Nx} = 1/2\Delta_x$ is typically referred to as the Nyquist frequency in light of the early contributions of Harry Nyquist [3].

Of course, the limits in (10.1) are not physically realizable, nor does a typical image have zero frequency content outside of the interval [0, f_{Nx}]. In fact, it can be shown that variations in image intensity occurring at frequencies outside of this interval will be falsely interpreted as occurring at frequencies inside the interval. This is a well-known phenomenon referred to as *aliasing* and if not corrected will manifest as spurious intensity levels at certain frequencies [2]. Figure 10.1 provides a pictorial description of the way in which frequency content beyond the Nyquist appears spuriously in the baseband. The MTF of the optical system depicted in the figure passes information at frequencies beyond Nyquist. The act of sampling the signal forces the higher frequency information to be "folded" into the Nyquist interval. The only way to eliminate aliasing artifacts is to filter spatial frequencies beyond f_{Nx} prior to sampling; however, this inevitably results in some loss of high-frequency image content.

For these reasons it turns out that the *sinc*() function is not ideal for interpolating images in practice, so researchers have turned to a variety of other functions. Many of these are given in [2] and include linear, cubic, or spline interpolation functions. In short, if we are given discrete samples of a band-limited image, we can always attempt to recover the intensity of that image at any spatial location using an appropriate interpolation function. Our success in this endeavor will depend on the image content (e.g., spatial frequencies), the interpixel spacing Δ_x, and the interpolation function used.

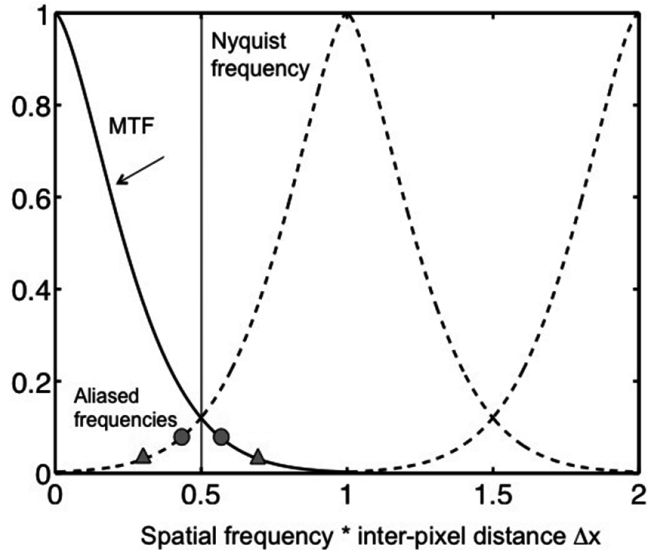


Figure 10.1 Information at frequencies beyond Nyquist is passed through the system MTF. However, this information is falsely "folded" into the sub-Nyquist band due to aliasing.

Of course, we seldom deal with one-dimensional images and the above-described procedure must be extended. The Nyquist frequencies in both the x and y spatial dimensions are given by $f_{Nx} = 1/2\Delta_x$ and $f_{Ny} = 1/2\Delta_y$, respectively, allowing for the accurate reconstruction of band-limited images with spatial frequency content below f_{Nx}, f_{Ny} cycle/mrad. As with the one-dimensional case, interpolation among the sampled points yields the reconstructed image at any spatial location. In general, given the sampled image $g(n\Delta_x, m\Delta_y)$ $n = 1 .. N$, $m = 1 .. M$ we can write

$$g_r(x,y) = \sum_n \sum_m g(n\Delta_x, m\Delta_y) h(x - n\Delta_x, y - n\Delta_y) \tag{10.2}$$

as a model for the recovered image, where $h()$ is the interpolating function. Common practice is to choose $h()$ from the family of separable functions; that is,

$$h(x,y) = h(x)h(y) \tag{10.3}$$

Thus, one may interpolate in the x and y directions independently using the functions $h(x)$,$h(y)$. For example, the cubic B-spline functions

$$h(x) = \begin{cases} (1/2)|x|^3 - |x|^2 + 2/3 & 0 \le |x| < 1 \\ -(1/6)|x|^3 + |x|^2 - 2|x| + 4/3 & 1 \le |x| < 2 \\ 0 & \text{otherwise} \end{cases} \tag{10.4}$$

can be used to interpolate among local regions of an image. Like most interpolation functions, $h(x)$ in this example has a small, finite support, extending only over 5 neighboring pixels. Thus, the spectral domain support is infinite and the filters are incapable of completely eliminating high-frequency image artifacts occurring due to effects such as aliasing. For these reasons it is often advantageous to design the interpolation filters in the frequency domain.

Note that mathematically, (10.2) is discretely convolving the sampled image with the interpolation function. Thus, we may have also written

$$G_r(f_x, f_y) = G(f_x, f_y) H(f_x, f_y) \tag{10.5}$$

where $H(f_x,f_y)$ describes the filter in the Fourier domain and can be constructed to guarantee finite support in the frequency domain and eliminate unwanted frequency content. Of course, this approach also comes with pitfalls, namely, that upon transforming back to the image domain $g_r(x,y)$ there will often be artifacts due to the inability of the Fourier series to accurately capture sharp frequency domain transitions. Whether or not one is willing to accept the spurious frequency content associated with spatial filters or the "ringing" that comes with Fourier domain filters depends on the application. In many cases it may be advantageous not to apply any filtering or interpolation and instead accept the aliased frequency content that will inevitably be present. A detailed discussion of these issues can be found in [4]. In this reference a number of linear filtering strategies are given for optimizing different aspects of image reconstruction.

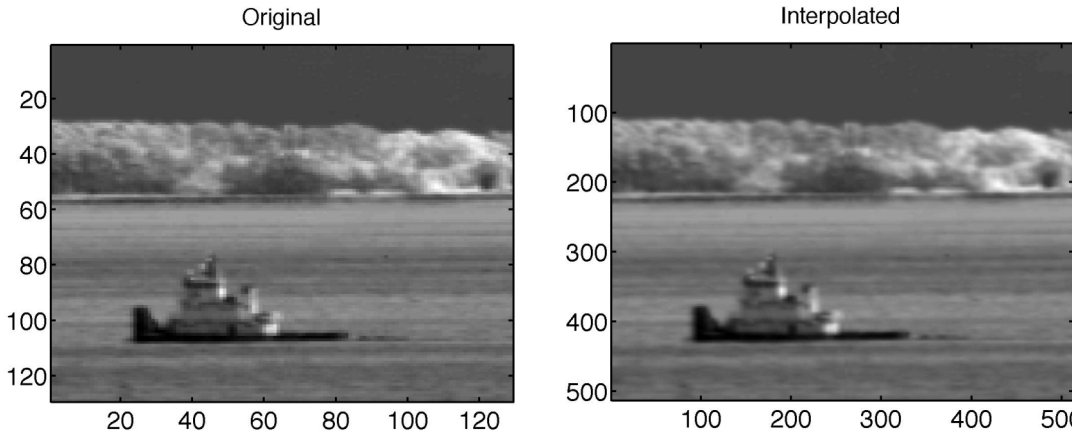


Figure 10.2 Low-pass filtering effect associated with the interpolation filter described by (10.4).

10.2 Applications of Image Filtering

As we have just shown, the process of interpolating among image points can be interpreted as a filtering operation. For example, the interpolation filters just described in (10.4) are low-pass filters that suppress high-frequency content in an image. Figure 10.2 shows the result of applying (10.4) to imagery from a short-wave infrared (SWIR) camera, resulting in a 4× improvement in resolution. However, as can be seen from the interpolated image, a "blurring" effect results from the suppression of the high-frequency content.

Although interpolation is inherently a low-pass operation, other types of filters can be designed for other purposes. Two such purposes are described in what follows.

Localized Contrast Enhancement

As the name implies, localized contrast enhancement is an operation designed to emphasize low-contrast areas in an image. The most basic contrast enhancement strategies are the so-called histogram equalization (HE) methods. These methods

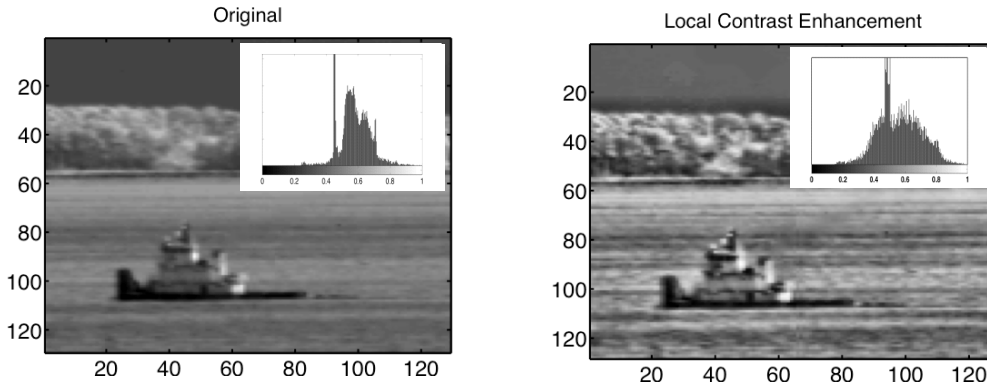


Figure 10.3 Localized contrast enhancement approach. The original image possesses a narrow pixel intensity histogram (inset), making areas of dissimilar intensity seem similar to the eye. The corrected image effectively spreads out the histogram allowing greater contrast in the resulting image.

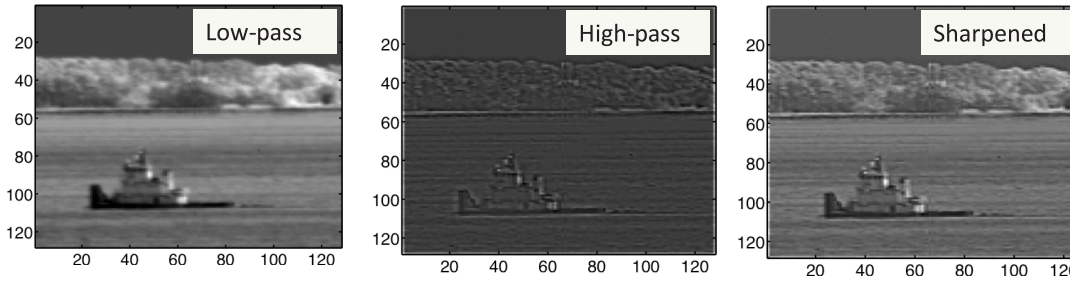


Figure 10.4 Unsharp-masking process. Image is first low-pass filtered using a 3×3 Gaussian blur kernel. This image is subtracted from the original (Figure 10.2) to create a high-pass copy of the image. Intensities above a certain threshold are amplified prior to adding back the low-pass image to yield the sharpened view.

transform the image pixel intensity values in a manner that amplifies regions of dissimilar intensity. Specifically, the transformation is designed such that the resulting histogram of intensity values is more uniformly distributed across the desired dynamic range (e.g., 0 to 255 for a grayscale image). The basic HE algorithm is a global approach whereby a single transformation is applied to the entire image. However, better results are often obtained by designing transformations that operate separately on local areas of an image. Figure 10.3 shows the result of applying a "localized" HE algorithm that performs the equalization on image segments (in this case 16- $\times$ 16-pixel image tiles) followed by the aforementioned bilinear interpolation to join the transformed tiles together [5–8]. Note that the original image possessed a near-Gaussian distributed set of intensity values with small variance (Figure 10.3, inset) resulting in a relatively low contrast. The local HE approach effectively spreads out this histogram such that areas of varying contrast are more clearly visible.

A number of other locally adaptive approaches have been derived. In nearly all of them the basic procedure is the same: Find a transformation (local or global) that maps similar intensity values to a broader range. Many of these transformations are designed specifically to amplify the intensity gradient associated with edges. The idea is to take areas of rapidly varying pixel intensity and stretch those areas to span a wider dynamic range. As such, these approaches can be appropriately viewed as high-pass filtering operations.

One popular class of approach is sometimes referred to as *unsharp-masking*. In this approach the connection to filtering is direct. The practitioner subtracts a low-pass filtered version of the image from the original, amplifies the residual high-frequency components (typically associated with edges), and then adds back in the low-pass image (see Figure 10.4). Methods differ mainly in the type of filter used [9, 10].

A related approach developed by Subr et al. [11] maximizes local image gradients (local defined by the nearest neighboring pixels) subject to constraints that bound the image to a prespecified range (e.g., 0 to 255). Edge enhancement can also include multiresolution approaches (e.g., wavelet, curvelet) whereby one decomposes the image, magnifies large coefficients (assumed to be those associated with edges), and then transforms back. Dippel et al. [12] compare both Laplacian pyramid and wavelet-based multiresolution techniques in terms of localized contrast

enhancement, whereas Starck et al. performed a multiresolution edge enhancement using the curvelet transform [13]. Each of these approaches also amplifies the high-frequency content associated with an edge and, hence, can be viewed as a form of high-pass filtering.

Boost Filtering

In certain instances we may wish to enhance certain spatial frequencies in an image at the expense of others. For example, the human eye is more sensitive to certain frequencies than others [14] and it is sometimes deemed advantageous to design a digital filter to "boost" or amplify these frequencies [15]. Denote the Fourier transform of an input image i at discrete frequencies f_k as $G_i(f_k)$. The filtered output image is given as

$$G_o(f_k) = G_i(f_k)H(f_k) \tag{10.6}$$

Allow the digital filter $H(f_k)$ to be the "N-tap" cosine filter, that is,

$$H(f_k) = \begin{cases} \sum_{n=0}^{(N-1)/2} A_n \cos(2\pi n f_k) & N \text{ odd} \\ \sum_{n=1}^{N/2} A_n \cos(\pi(2n-1)f_k) & N \text{ even} \end{cases} \tag{10.7}$$

where the coefficients A_n must sum to unity. This filter is shown graphically in Figure 10.5 for A = [0.76, 0.91, –0.21, –0.46] out to the (normalized) spatial Nyquist frequency.

Clearly frequencies in the range $0.1 < f_k < 0.2$ are being amplified and the output image will have greater intensity at these spatial frequencies. As before, the

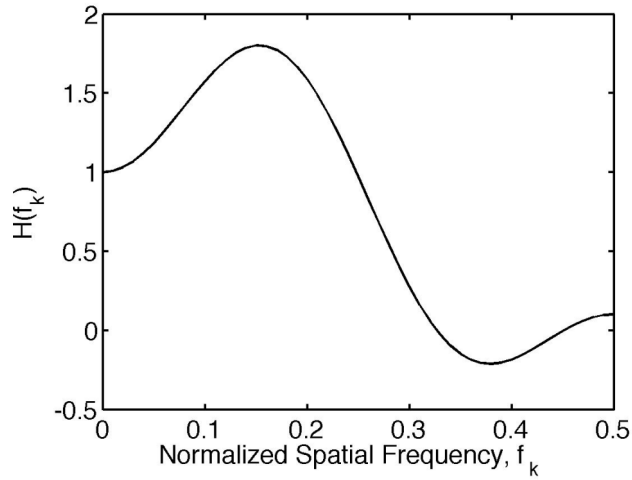


Figure 10.5 A digital boost filter showing amplification in the normalized spatial frequencies $0.1 < f_k < 0.2$.

extension of this 1D filter to two dimensions is accomplished by assuming separability such that the filtering is performed independently in both the x and y spatial directions. The overall effect on the output image can be tailored by varying the functional form of the filter and, of course, the filter coefficients. A more comprehensive look at this topic can be found in [16].

10.3 Super-Resolution Image Reconstruction

In general, when the diffraction spot size is smaller than the physical spacing between pixels, the associated imagery is said to be "undersampled" (i.e., the system MTF is passing frequencies the sampled data cannot resolve). As we have already pointed out, the focal plane cannot directly capture variations in image intensity occurring on spatial scales that are less than twice the pixel spacing, that is, the Nyquist limit. *Super-resolution* is a term used to describe a class of approaches for increasing this limit without changing the design of the optics and detectors, thereby preserving the high-frequency, detailed image information.

Super-resolution is typically cast as the linear inverse problem

$$\mathbf{g}_k = \mathbf{W}_k\mathbf{z} + \mathbf{n}_k \tag{10.8}$$

where the desired high-resolution image $\mathbf{z}$ must be inferred from a sequence of $k = 1 \ldots P$ low-resolution images that have been corrupted by additive noise $\mathbf{n}_k$. The matrices $\mathbf{W}_k$ relate the unknown high-resolution pixel values in $\mathbf{z}$ to the low-resolution values in each of the $\mathbf{g}_k$. This matrix is therefore a model that captures motion between frames ("warping"), system blur, and the downsampling operation, whereby the (unknown) high-resolution values are sampled by a low-resolution device [17]. Clearly if we have a single, low-resolution image, solving (10.8) is a highly underdetermined problem. However, *multiple* low-resolution images can be used to provide good estimates of $\mathbf{z}$.

An overview of the super-resolution image reconstruction algorithm is illustrated in Figure 10.6. The three major steps in super-resolution image reconstruction methods [2] are as follows:

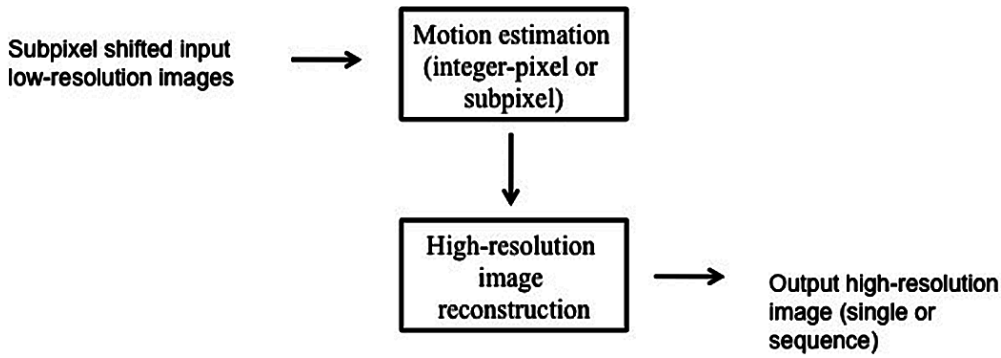


Figure 10.6 Overview of super-resolution image reconstruction.

1. Acquiring a sequence of $k = 1 \ldots P$ images from the same scene with subpixel shifts (fraction pixel displacements) among the images. The undersampled low-resolution images are captured either by natural jitter or some kind of controlled motion of the camera. In some cases, the images of successive frames contain not only subpixel shifts but also integer pixel shifts (gross shifts).
2. Motion estimation (integer pixel or subpixel):
 Integer pixel: The motion estimation algorithm estimates overall gross shifts of each frame with respect to a reference frame with an integer pixel accuracy. To compensate the integer pixel shifts among the input images, the algorithm realigns the input images.
 Subpixel: The motion estimation algorithm estimates the subpixel (fractional pixel) shifts for each frame with respect to a reference frame.
3. High-resolution image reconstruction. The reconstruction algorithm is applied to the low-resolution input images with the estimated subpixel shifts among images to obtain the high-resolution (alias-free) output [i.e., solve (10.8)]. The output is either a single high-resolution image that is generated from a collective sequence of low-resolution images or a sequence of high-resolution images, such as in a video sequence, that is generated from multiple sequences of low-resolution images.

A number of approaches can be used in each step. Here we provide a basic overview of these approaches along with appropriate references.

Image Acquisition: Microdither Scanner Versus Natural Jitter

The first step in super-resolution requires a sequence of images to be recorded with subpixel shifts between them. One approach is to have known, controlled pixel displacements applied by a specially designed sensor or scanner [18–20]. This method is sometimes referred to as *microscanning* or *mechanical dither*.

A second approach is to use natural (uncontrolled) motion to provide the subpixel shifts. Practically this is advantageous because in many applications the imager is mounted to a moving platform where mechanical vibrations are always present. In a rescue mission, for example, a helicopter, moving vehicle, or a moving ship may carry the camera. The drawback, of course, is that the unknown motion must be estimated prior to reconstruction.

Subpixel Shift Estimation

The next step in super-resolution is to estimate the subpixel shifts among a reference image and some number of additional images to be used in the reconstruction. Many shift estimation methods assume at the outset that the images to be matched differ only by an unknown translation, that is,

$$g_2(x, y) = g_1(x - x_s, y - y_s) \tag{10.9}$$

In practice, however, there may also be rotation, noise, and geometrical distortions; however, these are often not considered essential to the registration model. As a first step one typically computes a correlation function between two images, $g_1(x,y)$ and $g_2(x,y)$ as follows:

$$r(k,l) = \sum_{n=1}^{N} \sum_{m=1}^{M} R\left[g_1(n+k-1, m+l-1), g_2(n,m)\right] \tag{10.10}$$

where $1 \le \kappa \le N$, $1 \le l \le M$ and the function $R[\]$ denotes a similarity measure, such as a cross-correlation or absolute difference function. For appropriately chosen $R[\]$, the function $r(k,l)$ will be maximized by the unknown translation k,l. Of course, this approach can is only capable of capturing integer pixel value shifts, whereas what is sought are the subpixel distances x_s,y_s. Therefore, a common approach is to simply fit the function $r(k,l)$ to a quadratic or other "peaked" function, and take the maximum of this function to be the unknown shift between $g_1(x,y)$ and $g_2(x,y)$.

Other approaches include resampling via intensity domain interpolation and resampling via frequency domain interpolation. Both of these approaches are discussed in Young et al. [2]. We should mention, however, that the super-resolution overview of Park et al. [21] does not include such methods because they are incapable of recovering the high-frequency components that are lost due to the sampling process.

A different approach assumes a continuous, differentiable function relating the images $g_1(x,y)$ and $g_2(x,y)$. The job of the practitioner is to use this model to forecast the subpixel shift. Expanding (10.9) as a Taylor series gives the model

$$g_2(x,y) = g_1(x) + x_s \frac{\partial g_1(x,y)}{\partial x} + y_s \frac{\partial g_1(x,y)}{\partial y} \tag{10.11}$$

where x_s,y_s is the unknown shift we seek to estimate. Approximating the derivatives numerically yields 1 equation with two unknowns. However, this equation is assumed to hold across all pixels; thus, in practice, (10.11) represents a system of equations. Minimizing the mean square error between model and image data yields the desired estimates. The entire estimation procedure is described in [2].

A more sophisticated approach is the *optical flow method* whereby $g_1(x,y) = I(x,y,t_1)$ and $g_2(x,y) = I(x,y,t_2)$ represent the image intensities at two different times. The goal is to estimate the spatially dependent velocity vectors dx/dt and dy/dt that characterize the motion between images. Instead of assuming a smooth function connecting the images, the model assumes the image intensity at any point (pixel) is approximately constant from one time to the next, that is, $dI(x,y,t)/dt = 0$. Applying the chain rule yields the constraint equation

$$\frac{\partial I(x,y,t_1)}{\partial x}\frac{dx}{dt} + \frac{\partial I(x,y,t_2)}{\partial y}\frac{dy}{dt} + \frac{\partial I(x,y,t)}{\partial t} = 0 \tag{10.12}$$

As with the previous approach, we can approximate the derivatives numerically and solve (10.12) over all regions of the image for the unknown velocity vectors dx/dt and dy/dt. Of course, the solution to (10.12) requires a constraint, because there are two unknowns and a single equation. A commonly used constraint is to assume

the motion field varies smoothly in most of the image and that pixel-to-pixel variation of the velocity vectors can be quantified by the square of the velocity vector gradient. Thus, one finds the solution of (10.12) that minimizes

$$e_s^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \tag{10.13}$$

where $u = dx/dt$ and $v = dy/dt$ are the velocity vectors. Details of this approach can be found in Horn and Schunk [22] and Young et al. [2]. Once the velocity vectors have been estimated, they may be used to determine spatially dependent shifts in the image.

An example of optical flow–based motion estimation is shown in Figure 10.7. Clearly there is spatial variation in the estimated pixel shifts. This is in contrast to the correlation or gradient method, which produces a single shift, independent of location in the image.

Image Reconstruction

Given multiple low-resolution images and estimates of their respective degree of translation (more generally "warping") from a single reference image, the job of the reconstruction algorithm is to interpolate all subimages onto a single high-resolution, uniform grid of points. To this end, several algorithms are available to the practitioner.

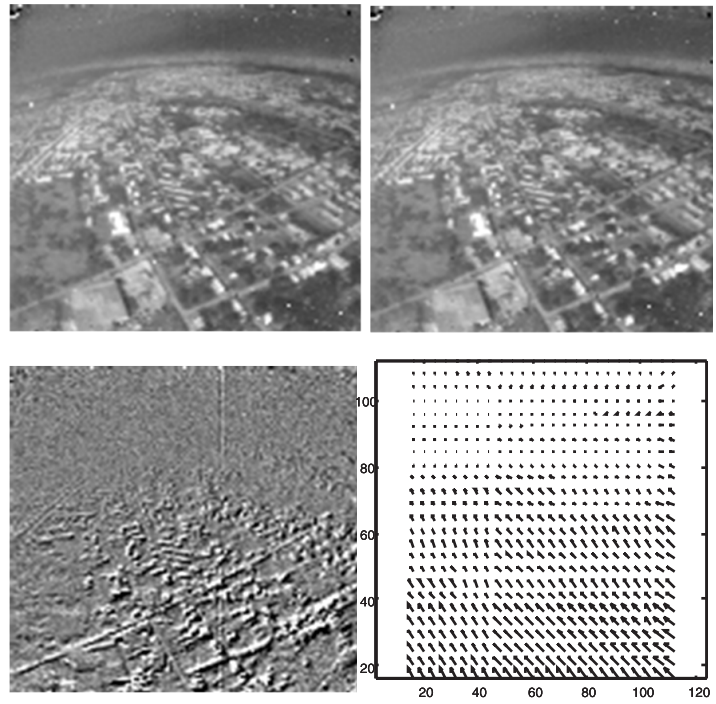


Figure 10.7 (Top) Two frames of aerial imagery, separated by 1/25th second, from an undersampled sensor. (Bottom left) The difference image between these two images. (Bottom right) Optical flow of this image sequence. Note the spatially varying optical flow, with the greatest flow occurring at points closest to the sensor. (From: [5].)

Nonuniform Interpolation

The most common situation is one in which the estimated subpixel shifts are arbitrary (due to random "dithering" of the low-resolution images) thus we require a method for relating nonuniformly spaced low-resolution samples to the high-resolution grid. To this end, various nonuniform interpolation approaches have been proposed, among them biharmonic spline interpolation [23], using Delaunay triangulation interpolation [24], using generalized multichannel sampling [25], using a warping procedure [26], using weighted interpolation [27], and convolving with a shifted-fractional kernel (polynomial kernel) [28]. The nonuniform approach is conceptually simple in that it takes the registered low-resolution images $\mathbf{g}_k(x,y)$ and directly interpolates to form the single, high-resolution image $z(x,y)$.

Regularized Reconstruction

Another approach is to solve (10.8) using a constrained least-squares approach. Specifically, one can choose the estimate to be that which minimizes

$$\hat{\mathbf{z}} = \min_{\mathbf{z}} \left[\sum_{k=1}^{P} \left\| \mathbf{g}_k - \mathbf{W}_k \mathbf{z} \right\|^2 + \alpha \left\| \mathbf{C}_k \mathbf{z} \right\|^2 \right] \tag{10.14}$$

The matrices $\mathbf{C}_k$ are high pass filters (e.g., derivative operator), thus α is a regularization term that penalizes nonsmooth images. The method is therefore predicated on the assumption that the image is smoothly varying spatially. Many solutions for (10.14) are available along with a number of different regularization strategies. Several of these are described in [4] [29–33].

Example and Performance Estimates

The most important question from the standpoint of designing optical systems is this: How do we quantify the performance gains offered by the above-described super-resolution methods? We begin by considering imagery collected from an airborne FLIR sensor and displayed in Figure 10.8(a). Infrared detectors, historically, are large (20 to 30 μm) compared to visible detectors (2 to 5 μm) so infrared systems (especially MWIR systems) are good candidates for seeing super-resolution performance increases. A subset of the raw imagery containing a ground vehicle is shown in Figure 10.8(b). At the sensor's natural resolution, it is extremely difficult to classify this or any other potential target of similar size. Figure 10.8(c) provides a super-resolved image, recovered from P = 16 low-resolution measurements that use the camera's natural jitter to provide the subpixel shifts. This reconstruction illustrates a significant improvement in the detailed information needed for classification (e.g., vehicle tires).

To demonstrate the improvement in target recognition, we summarize the results of a *triangle orientation discrimination* experiment, described in detail by Young et al. [2, Chp. 6]. Imagery of a triangle target, shown in Figure 10.9, is presented to an observer in one of four possible orientations.

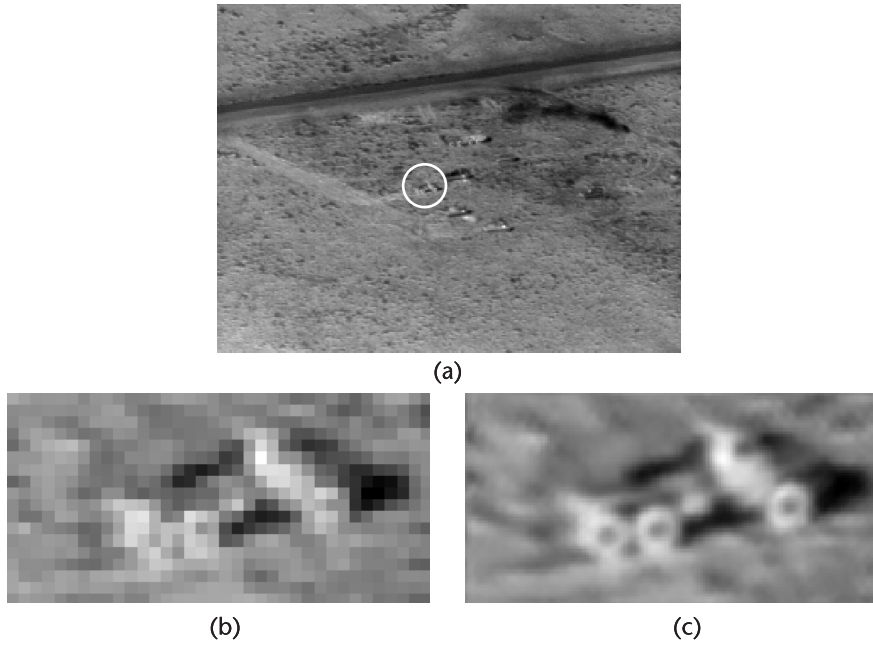


Figure 10.8 FLIR images of a ground vehicle from an airborne sensor: (a) entire scene, (b) one of the original low-resolution images, and (c) super-resolved image.

The observers are asked to classify the target orientation as a function of range using target imagery prepared in one of several ways. We first denote as "Algorithm_1" the method of Schuler et al. [34] whereby we compute the optical flow of the scene to calculate the scene motion, populate a high-resolution grid, and apply the interpolation algorithm. The other method by Young et al. [35], denoted "Algorithm_2," computes the subpixel shifts using the frequency domain correlation method and uses the error-energy reduction method to reconstruct the high-resolution image. In the rest of this section, we use the acronym *SR* for *super-resolution*. Referring to Figure 10.10:



Figure 10.9 Triangle target used in super-resolution experiments. Results are presented in Figure 10.10.

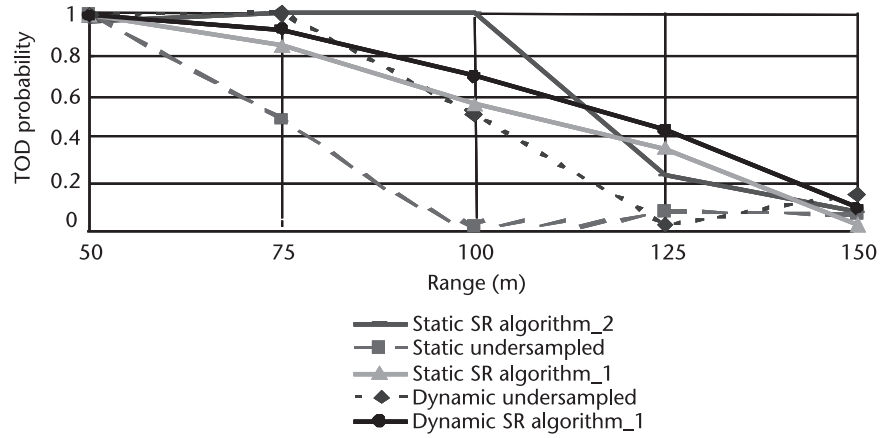


Figure 10.10 Experimental results associated with triangle orientation experiment. Super-resolution provides noticeable improvements in target classification.

1. *Static undersampled images*: These were taken by extracting a single frame from the image sequences.
2. *Static SR images*, *Algorithm*_1: These were image sequences that were processed with SR using Algorithm_1 to result in a single static super-resolved image.
3. *Static SR images*, *Algorithm*_2: These were image sequences that were processed with SR using Algorithm_2 to result in a single static super-resolved image.
4. *Dynamic undersampled images*: These were the raw image sequences (video) taken from the undersampled camera.
5. *Dynamic SR images*, *Algorithm*_1: These were image sequences that were processed with Algorithm_1 to result in a dynamic super-resolved image sequence.

All imagery was collected using an Indigo Systems Merlin LWIR uncooled microbolometer thermographic camera. The range varied from 50 m to 150 m in 25 m increments. Ranges were selected to vary observer performance, with and without processing, from very high probability to very low probability of a correct answer. The probability curves for the different cases are shown in Figure 10.10.

The baseline probability case is the static undersampled data, which resulted in poor range results with a 50% probability of target discrimination at a range of 75 m. The static SR Algorithm_1 and SR Algorithm_2 data should be compared with this baseline to determine the benefit of SR algorithms for static images that result from SR processing on sequences of undersampled imagery. Note that Algorithm_2 provides a high probability of target discrimination with a 50% probability out to 120 m, a significant improvement. Algorithm_1 provides a 50% probability of target discrimination at 110 m, also a significant improvement, but with a graceful degradation with range. For static imagery, the SR processing provides between 40% (Algorithm_1) to 60% (Algorithm_2) improvement in range performance.

The baseline for dynamic imagery is the dynamic undersampled data. When the dynamic SR Algorithm_1 data are compared with this baseline, the range at 50% target discrimination improves from 100 m to 120 m for a 20% increase in range performance. Although some improvement is realized, the effect is not on the scale of the improvement seen for the static case. Such comparisons suggest that the human eye is providing some degree of SR processing. That is, the eye–brain combination is putting together sequential images in some limited sense to accomplish a SR effect.

The concept of human perceptual SR is supported when comparing the two undersampled cases, the static versus dynamic. With no SR processing, the static undersampled range at 50% target discrimination is 75 m, whereas the dynamic undersampled range is 100 m, a 33% increase in range performance. The idea of human perceptual SR is also supported by the phenomenon of dynamic minimum resolvable temperature difference (MRTD), whereby an observer can see four-bar patterns when they are moving and cannot see them without motion [36].

10.4 Image Fusion

We conclude the chapter by discussing image fusion, the associated fusion algorithms, and metrics that can be used to assess the resulting image quality. Image fusion, as the name implies, is a method or technique for producing a single image from two or more images where the resulting fused image carries more information about the underlying scene. Although image fusion algorithms have been around for a number of years, the determination of imaging system performance with image fusion is relatively new. Interest is growing, however, in the ability to quantify the additional performance offered by more than one sensor and by the addition of image fusion to an imaging system.

Image fusion has a large number of applications and is used frequently when one sensor cannot satisfy the requirements associated with a specific system. For example, military helicopter operations are often constrained by the numerous environmental conditions associated with low-light levels and poor weather. Low-altitude operations are extremely difficult when flying over featureless terrain with low-scene contrast. In addition, low-altitude operations cause obscurations with wind-blown and recirculated dust that can cause a vision "brownout." These conditions yield poor situational awareness and can result in loss of aircraft control or, worse, in loss of crew and aircraft. Even at higher altitudes, fog, clouds, rain, and snow can cause loss of image contrast and result in hazardous situations. In this particular application, image fusion has been used successfully to combine infrared imagers and low-light-level television sensors. The benefits of fusion have been demonstrated by flight trials conducted in these degraded environmental conditions [2].

Fusion Algorithms

The objective of fusion is to combine the relevant information from two or more source images to form a composite image of higher information content and fidelity. This must be done in such a way as to minimize artifacts associated with the fusion algorithm (misregistration, blurring, and so forth).

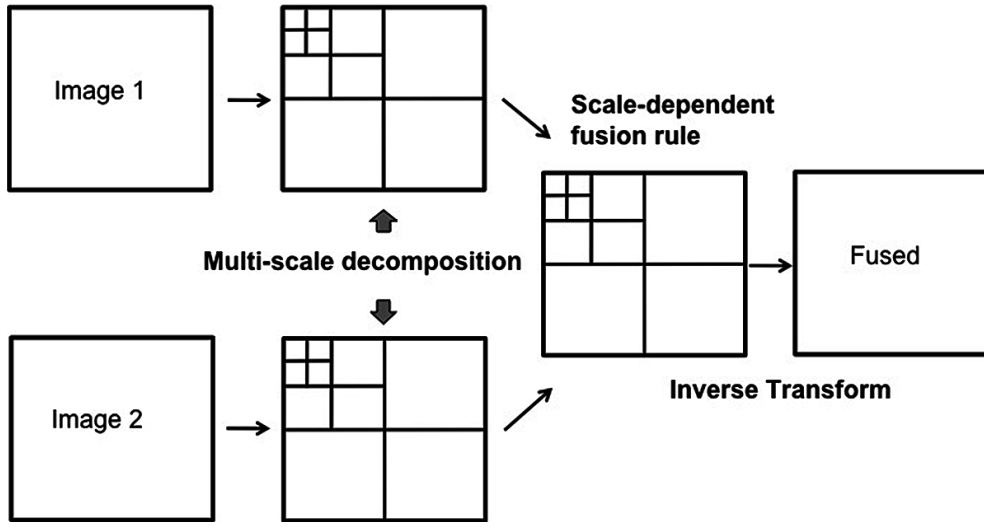


Figure 10.11 Schematic of multiresolution image fusion.

The simplest fusion algorithms are those that operate at the pixel level, resulting in pixel-level fused image output. For example, the basic superposition algorithm takes two images $A(x,y)$ and $B(x,y)$ and forms the fused image

$$I(x,y) = \alpha A(x,y) + \beta B(x,y) \tag{10.15}$$

where x,y denote pixel location and α,β are nonnegative scalars that sum to 1. Setting both to 0.5 produces a simple average of the individual images. A more sophisticated approach is to perform a multiresolution decomposition of the images, combine the imagery using a scale dependent rule, and then perform the inverse transformation. This general prescription is shown schematically in Figure 10.11.

Laplacian Pyramid

One of the earliest approaches to multiscale analysis is the so-called Laplacian pyramid, described in [37]. Each of the images to be fused is processed by a succession $l = 0 \ldots N$ of low-pass filters, producing blurred images

$$I_l(x,y) = \sum_{m=-M_l}^{M_l} \sum_{n=-M_l}^{M_l} w(m,n) I_{l-1}(2x+m, 2y+n) \tag{10.16}$$

where the original image is taken as I_0. The filters are defined by the weights $w(m,n)$ and are assumed zero outside of the user-defined support $-M_l \ldots M_l$. Thus, at each level l both the number of pixels and (by construction) filter bandwidth are reduced by a factor of 2. Typical implementation sets $M_l = 2$ and assumes the kernel is separable, that is, $w(m,n) = w(m)w(n)$. The standard implementation assigns the 1D weights based on a Gaussian shape function [35]; however, other shapes may also be used [37].

Rather than work directly with this new image set, Burt and Adelson showed that a better encoding strategy is to use the *difference* between successive levels [37]

(the differencing operation removes interpixel correlations and, hence, provides better compression). The problem is that one cannot directly difference images of two different sizes. Burt and Adelson therefore define an interpolation function *EXPAND*{} as follows:

$$EXPAND\{I_{l+1}(x,y)\} \equiv 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m,n)I_{l+1}\left(\frac{x+m}{2},\frac{y+n}{2}\right) \tag{10.17}$$

where the arguments $(x + m)/2$ and $(y + n)/2$ contribute only when they are integers. We can now define the Laplacian pyramid for level l as

$$L_l(x,y) = I_l(x,y) - EXPAND\{I_{l+1}(x,y)\} \tag{10.18}$$

The sequence of filtered image differences provides a multiresolution depiction of the original image.

Returning to the fusion problem, the goal is to combine the decompositions of two different images in a manner that enhances relevant features in each of the constituent images. Denoting $L_l^{(1)}(x,y)$ and $L_l^{(2)}(x,y)$ as the Laplacian pyramids associated with images $I^{(1)}(x,y)$, $I^{(2)}(x,y)$ one may define a fusion rule, for example,

$$L_l^{(1,2)}(x,y) = \left\{\begin{array}{ll} L_1^{(1)}(x,y) \text{ if } & \left|L_1^{(1)}(x,y)\right| > \left|L_1^{(2)}(x,y)\right| \\ L_2^{(2)}(x,y) & \text{otherwise} \end{array}\right\} \tag{10.19}$$

and then reconstruct the fused image starting with $I_N(x,y) = L_N(x,y)$ and forming, for $l = N - 1 \ldots 0$,

$$I_l^{(1,2)}(x,y) = L_l^{(1,2)} + EXPAND\{I_{l+1}^{(1,2)}(x,y)\} \tag{10.20}$$

The idea, of course, is to take the most prominent scale-dependent features from the constituent images and combine them in the fused image. This general approach stays the same, regardless of the specific decomposition used. For example, Toet used a modified version of the Laplacian pyramid for image fusion based on the ratios of an image pyramid [38].

As an example of Laplacian pyramid fusion, consider both a SWIR and longwave infrared (LWIR) image of an HMMV ground vehicle as shown in Figures 10.12(a) and (b), respectively. Applying the Laplacian pyramid algorithm as described above results in the fused image of Figure 10.12(c). The driver of the vehicle is only visible in the SWIR while the soldier in the trees can only be detected in the LWIR. The fused imagery clearly shows both.

Wavelet-Based Methods

More recently, wavelets have become the preferred signal model for multiresolution analysis. The discrete wavelet transform offers a wide array of filters for decomposing a signal into multiple scales. Furthermore, the algorithms for implementing the wavelet transform are fast and widely available. In the context of fusion, one can proceed as described next.

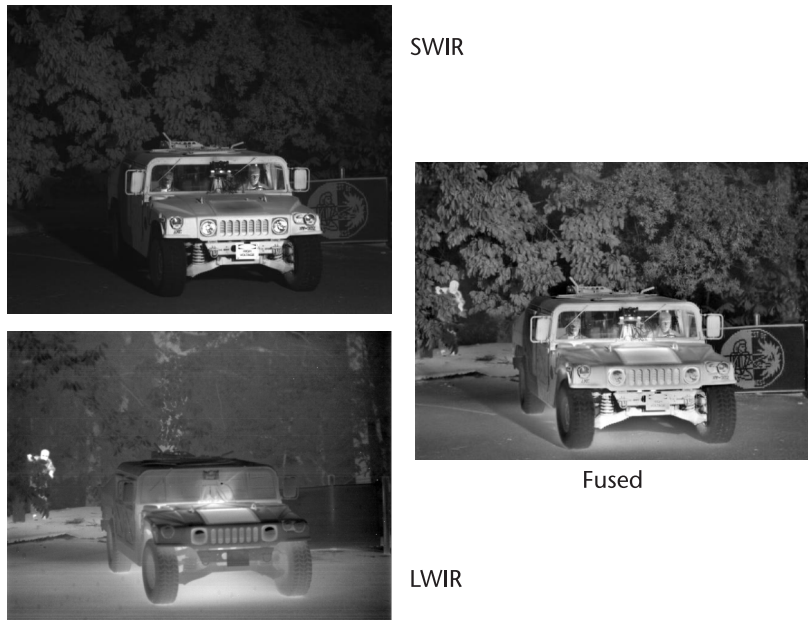


Figure 10.12 Laplacian pyramid example: (a) SWIR, (b) LWIR, and (c) fused image. (Images courtesy of Mike Prairie and Roy Littleton, U.S. Army Night Vision and Electronic Sensors Directorate.)

Huntsberger and Jawerth [39] describe standard wavelet decomposition of a signal, f, as a linear combination of translations and dilations of a single function ψ:

$$f(x) = \sum_{v} \sum_{k} C_{v,k} \psi_{v,k}(x) \tag{10.21}$$

where

$$\psi_{v,k}(x) = 2^{v/2} \psi(2^{v} x - k) \tag{10.22}$$

for integers v and k. The process of mapping $f(x)$ into the coefficients $C_{v,k}$ is called the wavelet transform of $f(x)$. The function $\psi_{v,k}$ is varied on a scale of 2^{-v} at the location $2^{-v}k$ and the corresponding coefficient $C_{v,k}$ depends of the function $f(x)$ at that scale and location.

Wavelet transform sensor fusion is accomplished in the same basic fashion as Laplacian fusion: by combining the wavelet coefficients of the two source images as in (10.18) to form a fused wavelet description. The inverse wavelet transform then recovers the fused image [39, 40].

The discrete wavelet transform is not shift invariant since the image is downsampled (each wavelet series image is downsampled by a factor of 4). That is, the output depends on shifts in the image so artifacts can be present when two images are fused that are not perfectly aligned (registered). The shift-invariant discrete wavelet transform (SIDWT) is a solution to the shift-variance problem, where the downsampling operation is removed from the wavelet decomposition process. However, because the images are not reduced in size, the operation is very inefficient since the number of calculations grows significantly compared to the

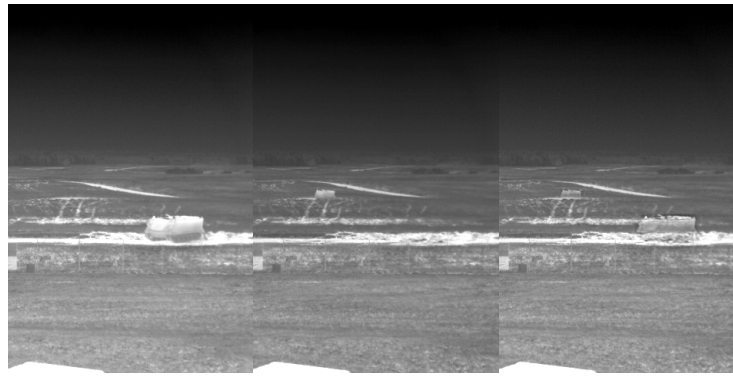


Image 1 Image 2 Fused image

Figure 10.13 Discrete wavelet transform example. Image 1 has a target in the foreground. Image 2 has a target in the background. The fused image shows both targets. (Images courtesy of Jonathan Fanning, U.S. Army Night Vision and Electronic Sensors Directorate].

DWT. The SIDWT is very applicable to component images that are not exactly registered.

Figure 10.13 shows a DWT fusion example. Image 1 shows a scene with a target in the foreground and image 2 shows the same scene, but with the target in the background. The DWT fusion provides an image where both target positions appear in the output image.

In short, the field of image fusion holds great promise. As more and more tactical imaging systems are equipped with multicolor (e.g., visible and SWIR) sensors, the full benefits of image fusion can begin to be realized. Researchers have begun to focus fused imaging systems on particular applications. For example, the work of Toet [41] fused infrared and visible imagery for the purpose of target detection, while Mull et al. [42] used MWIR and LWIR imagery to improve target discrimination. The field is also evolving with respect to quantifying fused image performance, a necessary step before widespread implementation. To this end, *image fusion quality indices* have been created to reflect different goals of fusion and include the metrics put forth by Piella and Heijmas [43], Xydeas and Petrovic [44], and Wang and Bovik [45].

10.5 Summary

The field of image processing comprises far too many topics to be covered in a single chapter. We therefore focused first on the basics of image acquisition (sampling), interpolation, and filtering and then described two approaches that we feel show great promise in detection and classification applications. Super-resolution is designed to overcome the inherent physical limitations of sampled systems through clever processing of multiple images. We have described the basic ingredients of super-resolution approaches and shown the benefits of super-resolved imagery. We have also described methods for image fusion whereby multiple sensing modalities can be combined to present the viewer with more informative imag-

ery. Examples of both the Laplacian pyramid and wavelet-based fusion techniques were provided.

References

[1] Shannon, C. E., "Communication in the presence of noise," *Proceedings of the IRE 37(1)*, 10–21, 1949, Reprinted in *Proc. IEEE*, 86(2), 1998.

[2] Young, S. S., R. G. Driggers, and E. L. Jacobs, *Signal Processing and Performance Analysis for Imaging Systems*, Norwood, MA: Artech House, 2008.

[3] Nyquist, H., "Certain topics in telegraph transmission theory," *Trans. AIEE*, Vol. 47, pp. 617–644, 1928, Reprinted in *Proc. IEEE*, 90(2), 2002.

[4] Vollmerhausen, R. H., and R. G. Driggers, *Analysis of Sampled Imaging Systems*, Bellingham, WA: SPIE Press, 2000.

[5] Pizer, S. M., E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, et al., "Adaptive Histogram Equalization and its Variations," *Computer Vision, Graphics and Image Processing*, 39, 1987, pp. 355–368.

[6] Zuiderveld, K., "Contrast Limited Adaptive Histogram Equalization" *Graphics Gems IV*, Cambridge, MA" Academic Press, 1994, pp. 474–485.

[7] Srinivasan, S., and N. Balram, "Adaptive Contrast Enhancement Using Local Region Stretching," *Proceedings of ASID'06*, New Delhi, India, 2006, pp. 152–155.

[8] Cheng, H. D., and X. J. Shi, "A Simple and Effective Histogram Equalization Approach to Image Enhancement," *Digital Signal Processing*, Atlanta, GA, Vol. 14, 2004, pp. 158–170.

[9] Arici, T., and Y. Altunbasak, "Image Local Contrast Enhancement Using Adaptive Nonlinear Filters," *Proc. IEEE Int. Conf. on Image Processing (ICIP 2006)*, 2006, pp. 2881–2884.

[10] Ramponi, G., "A Cubic Unsharp Masking Technique for Contrast Enhancement," *Signal Processing*, 67(2), 1998, pp. 211–222.

[11] Subr, K., A. Majumder, and S. Irani, "Greedy Algorithm for Local Contrast Enhancement of Images," in *Image Analysis and Processing – ICIAP* 2005, pp. 171–179, F. Roli and S. Vitulano (eds), Springer-Verlag, 2005.

[12] Dippel, S., M. Stahl, R. Wiemker, and T. Blaffert, "Multiscale Contrast Enhancement for Radiographies: Laplacian Pyramid Versus Fast Wavelet Transform." *IEEE Trans. Med. Imaging*, Cagliari, Italy, 2002, pp. 343-353.

[13] Starck, J.-L., F. Murtagh, E. J. Candes, and D. L. Donoho, "Gray and Color Image Contrast Enhancement by the Curvelet Transform," *IEEE Transactions on Image Processing*, 12(6), 2003.

[14] Granger, E. M., and K. N. Cupery, "An Optical Merit Function (SQF) Which Correlates With Subjective Image Judgments," *Photographic Science and Engineering*, 16, 1972, pp. 221–230.

[15] Holst, G. C., "Electro-optical Imaging System Performance," *SPIE Press*, 1995.

[16] Gonzalez, R. C., and E. Woods, *Digital Image Processing, Third Edition*, Upper Saddle River, NJ: Prentice Hall, 2008.

[17] Elad, M., and A. Feuer, "Restoration of a Single Superresolution Image From Several Blurred, Noisy, and Undersampled Measured Images", *IEEE Transactions on Image Processing*, 6(12), pp. 1646–1658, 1997.

[18] Zalevsky, Z., N. Shamir, and D. Mendlovic, "Geometrical superresolution in infrared sensor: experimental verification," *Optical Engineering*, 43(6), 2004, pp. 1401–1406.

[19] Ben-Ezra, M., A. Zomet, and S. K. Nayar, "Video super-resolution using controlled subpixel detector shifts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, 2005, pp. 977–987.

[20] Krapels, K., R. Driggers, R. Vollmerhausen, and C. Harford, "Performance comparison of rectangular (4-point) and diagonal (2-point) dither in undersampled IRFPA imagers," *Applied Optics*, Vol. 40, No. 1, 2001, pp. 101–112.

[21] Park, S. C., M. K. Park, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine*, 2003, pp. 21–36.

[22] Horn, B. K. P., and B. G. Schunk, "Determining optical flow," *Artificial Intelligence*, 17: 185–203, 1981.

[23] Bose, N. K., "Superresolution from image sequence", *Proceedings of IEEE International Conference on Image Processing 2004*, 2004, pp. 81–86.

[24] Lertrattanapanich, S., and N. K. Bose, "High resolution image formation from low resolution frams using Delaunay Triangulation," *IEEE Transactions on Image Processing*, Vol. 11, No. 12, 2002, pp. 1427–1441.

[25] Ur, H., and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graphical Models and Image Processing*, Vol. 54, 1992, pp. 181–186.

[26] Chiang, M. C., and T. E. Boult, "Efficient super-resolution via image warping," *Image and Vision Computing*, Vol. 18, No. 10, 2000, pp. 761–771.

[27] Alam, M. S., J. G. Bognar, R. C. Hardie, and B. J. Yasuda, "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," *IEEE. Trans. Instrumentation and Measurement*, Vol. 49, No. 5, 2000, pp. 915–923.

[28] Candocia, F. M., and J. C. Principe, "Super-resolution of images based on local correlations," *IEEE Trans. on Neural Networks*, Vol. 10, No. 2, 1999, pp. 372–380.

[29] Schulz, R. R., and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, Vol. 5, 1996, pp. 996–1011.

[30] Hardie, R. C., K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, Vol. 6, 1998, pp. 1621–1633.

[31] Elad, M., and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Processing*, Vol. 10, No. 8, 2001, pp. 1187–1193.

[32] Farsiu, S., M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Processing*, Vol. 13, No. 10, 2004, pp. 1327–1344.

[33] Lee, E. S., and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Trans. Image Processing*, Vol. 12, No. 7, 2003, pp. 826–837.

[34] Schuler, J. M., J. G. Howard, P. Warren, and D. Scribner, "TARID-based image superresolution," *Proceedings of SPIE Infrared and passive millimeter-wave imaging systems; design*, *analysis*, *modeling*, *and testing*, Vol. 4719, Orlando, FL, 2002, pp. 247–254.

[35] Young, S. S., and R. G. Driggers, "Super-resolution image reconstruction from a sequence of aliased imagery," *Applied Optics*, Vol. 45, No. 21, 2006.

[36] Webb, C., and C. Harford, "Dynamic minimum resolvable temperature difference testing for staring array imagers," *Opt. Eng.* Vol. 38, No. 5, 1999, pp. 845–851.

[37] Burt, P. J., and E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code," *IEEE Transactions on Communications*, 31(4), pp. 532–540.

[38] Toet, A., "Image fusion by a ratio of low-pass pyramid," *Pattern Recognition Letters*, Vol. 9, 1989, pp. 245–253.

[39] Huntsberger, T., and B. Jawerth, "Wavelet based Sensor Fusion," *Proceedings of SPIE*, Vol. 2059, 1993.

[40] Zheng, Y., E. Essock, and B. Hansen, "Advanced Discrete Wavelet Transform Fusion Algorithm and Its Optimization by Using The Metric of Image Quality Index," *Optical Engineering*, Vol. 44(3), 2005, pp. 037003.

[41] Toet, A., "Color Image Fusion For Concealed Weapon Detection," *Proceedings of SPIE*, Vol. 5071, 2003.

[42] Muller, M., O. Schreer, and M. Lopez Saenz, "Real-Time Processing and Fusion for a New High Speed Dual Band Infrared Camera," *Proc. SPIE*, Vol. 6559, 2007.

[43] Piella, G., and H. Heijmans, "A New Quality Metric for Image Fusion," *International Conference on Image Processing*, Barcelona, Spain, 2003, pp. 173–176.

[44] Petrovic, V., and C. Xydeas, "Sensor Noise Effects on Signal-Level Image Fusion Performance," *International Conference on Image Processing*, Barcelona, Spain, 2003.

[45] Wang, Z., and A. Bovik, "A Universal Image Quality Index," *Signal Processing Letters, IEEE*, Vol. 9(3), 2002, pp. 81–84.

CHAPTER 11

Displays, Human Perception, and Automatic Target Recognizers

Displays are the interface between the I^2R or EO sensor and the human vision system (Figure 11.1). The display converts the electrical signal from the I^2R or EO sensor to a visible signal that is subsequently presented to the viewer. Occasionally, systems are designed where humans do not interpret the data, so displays are not required. Imaging missile seekers with automatic target trackers are a good example. Various levels of human interpretation of imagery are seen in the targeting community. One example that is increasingly successful and gaining popularity is the targeting sensor coupled to a computerized automatic target recognizer (ATR), automatic target cueing (ATC) system, or aided target recognizer (AiTR). We collectively refer to these automated target exploitation systems as ATRs.

The performance of displays can be characterized with a modulation transfer function, where the MTF depends on the resolution of the display. The human visual system also has a transfer function that is based on the resolution limits of the eye. Both of these transfer functions are discussed; however, this chapter provides only an introduction on how to account for these "system" components.

Automatic targeting systems are in still early development, so the performance of these systems is difficult to quantify. An introduction to the ATR nomenclature is provided. ATR systems are typically nonlinear and do not lend themselves to linear systems modeling.

11.1 Displays

Much work is currently being performed on the analysis of displays. The evaluation of displays is still highly subjective because the primary basis of display design is viewer satisfaction. However, an increasing amount of work is being done in the objective performance evaluation of these devices. Pinson [1] states that display criteria are usually given in terms of the probability that the viewer can perform specific tasks. These tasks involve extracting information from a displayed image. The probabilities are related to more trackable criteria, such as measurements of MTF, contrast, SNR, and resolution.

The magnification of a system depends on the sensor's imaging characteristics and the observer/display geometry. The definition of system magnification (which is different from the optics magnification) is the ratio of the angular image size seen by the observer to the angular size of the image as seen by the sensor input pupil (see Figure 11.2):

$$M = \theta_{image} / \theta_{object} \tag{11.1}$$

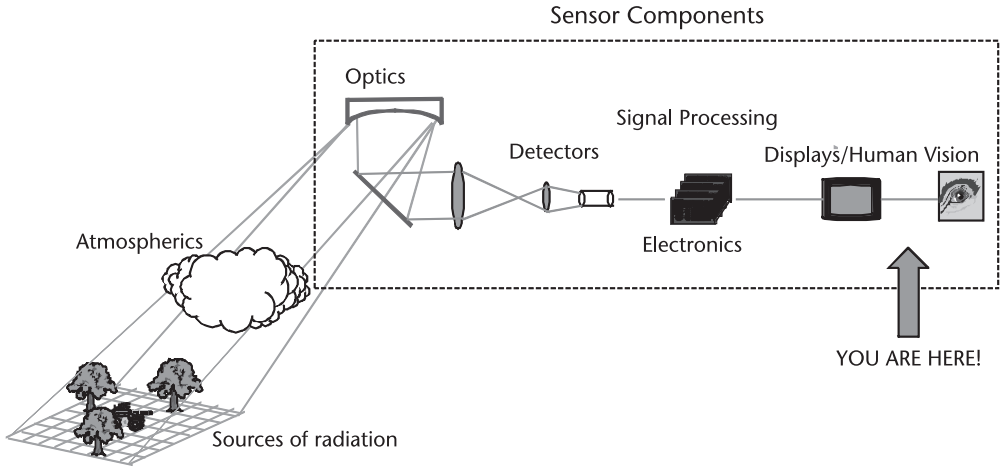


Figure 11.1 Systems Display/Human Vision Component.

This can be simplified so that the system magnification is not object dependent by writing

$$M = FOV_d/FOV_v \tag{11.2}$$

where FOV_d is the display field of view in the vertical direction and FOV_v is the sensor input pupil vertical field of view. Note that the magnification is larger if the viewer moves closer to the display. However, the image becomes objectionable because the near-point limitation of the eye and the cognitive frame of the observer. Viewing distances are typically selected by observers to make the image display lines unnoticeable. In aviation systems, typical viewing distances vary from 20 to 28 in., depending on the height of the pilot.

System displays include cathode-ray tubes (CRTs), light-emitting diodes (LEDs), and liquid-crystal displays (LCDs). These display systems are described in the following sections along with their corresponding resolutions and MTFs. Keep in mind that a large number of display types, each with its own MTF, must be accounted for in the wide variety of sensor systems. The ones covered here are some of the more common displays.

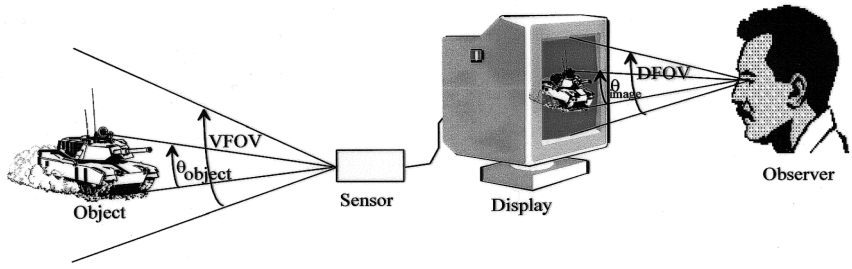


Figure 11.2 System magnification.

Example 11.1
A 3- by 4-deg sensor has an output image on a 15-cm-high display monitor. The observer is positioned 38 cm from the monitor. Determine the magnification of the system. The solution is given by (11.2) where the vertical FOV of the sensor is 3 deg. Typically, the horizontal FOV is larger than the vertical FOV and it is customary to specify vertical and then horizontal FOVs. The display vertical FOV is

$$FOV_d = tan^{-1}\left(\frac{15}{38}\right) = 21.5 \quad [\text{deg}]$$

The magnification is then 21.5/3 = 7.2. Note that a smaller sensor FOV or a larger display FOV gives a larger magnification.

11.2 Cathode-Ray Tubes

CRTs used to be the most common display system. In the first edition of this book, that was certainly the case. A CRT comprises an evacuated tube with a phosphor screen. An electron beam is scanned across the phosphor screen in a raster pattern as shown in Figure 11.3. The beam direction is controlled with horizontal and vertical magnetic fields that bend the beam in the appropriate direction. The phosphor converts the electron beam into a visible point, where the visible luminance of the phosphor is related to the electron beam current (and voltage). The standard raster scan and interlace pattern are also shown in Figure 11.3. First the solid line shows a field pattern that is traced out on the screen. The dashed line shows a second field pattern that is interlaced between the first field lines. Both fields together provide 525 lines, where around 490 of the lines display information. The two fields in the standard 2:1 interlace pattern are collectively called a *frame*. The fields are presented at a rate of 60 Hz and frames are presented at 30 Hz.

The output luminance of the phosphor spot is related to the tube voltage in the linear region (i.e., the region that is not saturated) by

$$L = k(V - V_T)^{\gamma} \tag{11.3}$$

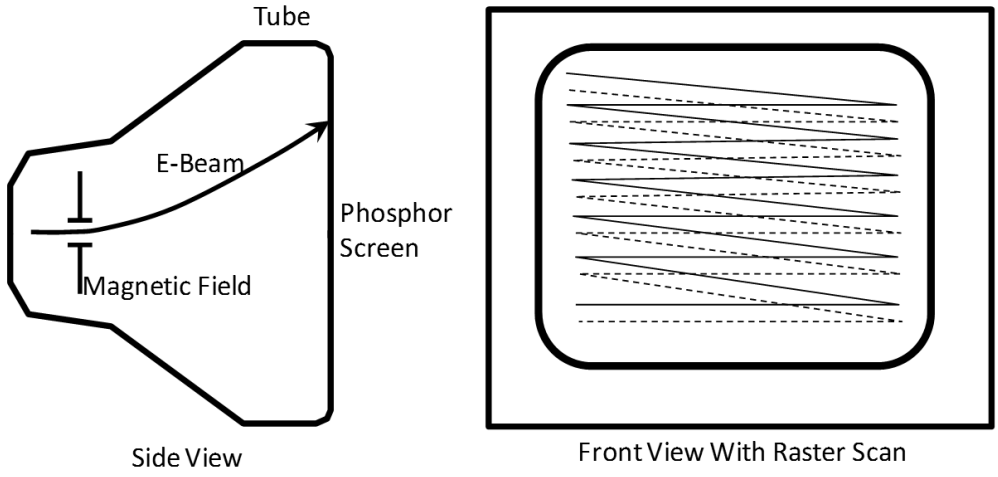


Figure 11.3 Cathode-ray tube.

where k is a constant, V_T is a threshold voltage, and γ is a display constant known as the display gamma. A high gamma provides more phosphor contrast between two different voltages.

A common CRT resolution assumption is that the electron beam spot and the corresponding luminescent spot are Gaussian in shape. The spot can be approximated by

$$S(x,y) = \frac{1}{2\pi\sigma^2} Gaus\left(\frac{\sqrt{x^2+y^2}}{\sqrt{2\pi}\sigma}\right) \tag{11.4}$$

where x and y are the horizontal and vertical positions on the screen and σ describes the size of the spot (known as the spot size parameter). Equation (11.4) can be considered the impulse response of the CRT. The MTF is then determined by taking the Fourier transform of this expression, yielding

$$MTF_{CRT}(\xi,\eta) = Gaus(\sqrt{2\pi}\sigma\sqrt{\xi^2\eta^2}) \tag{11.5}$$

Gaussian functions are separable in x and y so that the MTF given in (11.5) can be separated to give an MTF in the horizontal and vertical directions:

$$MTF_{CRT}(\xi,\eta) = MTF_{CRT}(\xi)MTF_{CRT}(\eta) = Gaus\left(\sqrt{2\pi}\sigma\xi\right)Gaus\left(\sqrt{2\pi}\sigma\eta\right) \tag{11.6}$$

The spot size can be measured using the *shrinking raster* technique [2]. With this technique, the line spacing on the CRT is reduced until the observer can no longer resolve the scan lines. The distance between the lines l can be estimated by measuring the total height of all lines and dividing by the number of lines. The spot-size parameter σ is related to the distance between the lines by

$$\sigma = 0.54l \quad [\text{m}] \tag{11.7}$$

Note that σ is in meters, so the MTF given in the above equations has an argument of cycles per meter. To provide an MTF that can be used in an angular system MTF estimate, the argument must be given in terms of cycles per radian or cycles per milliradian. This conversion is determined with the sensor FOV in milliradians and the size of the CRT display in meters. An example of this conversion in the vertical direction is

$$\sigma_{angular} = \sigma\frac{FOV_V}{MonitorSize_V} \tag{11.8}$$

where FOV_V is in milliradians. This angular spot-size parameter provides an MTF in the same spatial frequency domain as that of the other sensor components.

Example 11.2

The spot-size parameter σ has been determined to be 200 μm on a 15-cm-high CRT monitor using the shrinking raster technique. The monitor displays the output of a 5- by 7- deg FOV sensor. Plot the MTF of the monitor for system-related object space in units of cycles per milliradian.

The solution is given by first converting the spot-size parameter from micrometers to milliradians:

$$\sigma_{angular} = 200 \times 10^{-6}\,\text{m}\,\frac{87.3\ \text{mrad}}{15 \times 10^{-2}\,\text{m}} = 0.116\ \text{mrad}$$

This angular spot size can be used in the horizontal component of (11.6) to give the horizontal MTF. The MTF is given by

$$Gaus\left(\sqrt{2\pi}\,\sigma_{angular}\,\xi\right) = e^{-\pi\left(\sqrt{2\pi}\sigma_{angular}\xi\right)^2} = e^{-2\pi^2\sigma_{angular}^2\xi^2}$$

which is shown in Figure 11.4. If the *Gaus* function is written in terms of $Gaus\left(\frac{\xi}{\xi_{co}}\right)$ where ξ_{co} is the cutoff frequency, then it can be shown that $\xi_{co} = 1/\left(\sqrt{2\pi}\sigma_{angular}\right)$. In this case, the cutoff frequency is 3.4 cycle/mrad.

The CRT usually has a sample-and-hold circuit that gives a discrete number of horizontal pixels as the display spot scans the screen in the horizontal direction. This circuit samples the horizontal signal on a scan line into an integer number of samples and then holds the phosphor spot voltage constant over this sample as the spot is scanned across the horizontal direction. For example, many CRTs have 480 active lines in the vertical direction (out of the 525 possible lines). With a 1.33 horizontal-to-vertical image aspect ratio, the horizontal scan across the CRT screen would be broken into 640 samples. If the CRT has this type of sample-and-hold circuit, the psf for the sample-and-hold circuit is $h(x,y) = (640/FOV_h)rect\left[\frac{x}{FOV_h/640}\right]\delta(y)$ and the display sample-and-hold transfer function is $H(\xi,\eta) = sinc[(FOV_h/640)\xi]$. This transfer function must be accounted for in any system that includes a CRT display sample-and-hold circuit.

11.3 Light-Emitting Diodes

LEDs are P-N junction devices that emit light when they are forward biased. Because LEDs [3] emit light, they are very visible in dark conditions and are less visible

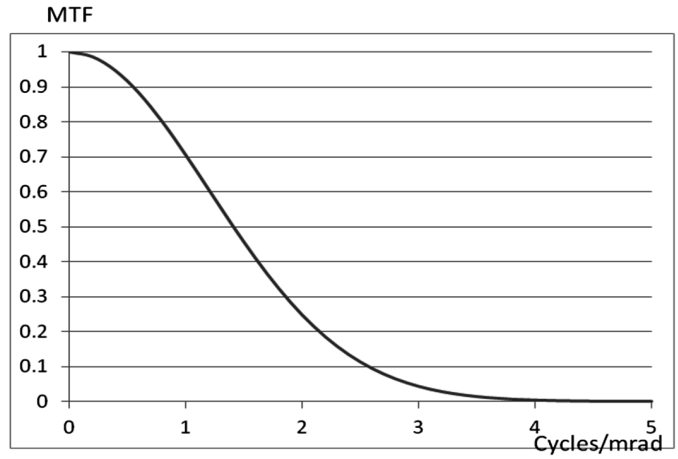


Figure 11.4 CRT example results.

in bright conditions. LEDs are luminescent (self-emitting in visible wavelengths), where the luminescence is proportional to the current passing through the LED. The visibility of the LED depends on the emitted radiation and the external quantum efficiency of the eye at the LED emission wavelength. The efficiency of LEDs is limited because of the mismatch of refractive index between air and the luminescent material (index of refraction around 3.6). Therefore, the emitted angle for total internal reflection is small. A lens can increase the efficiency by a factor of 2 to 3. LEDs today are used in large outdoor displays on a discrete level and on flat-panel displays indoor. Not many LED flat-panel displays are in current use in military systems. However, there a numerous older systems that used linear LED scanned arrays as described in the following paragraphs.

In past infrared imaging applications, LEDs were typically arranged in a linear array (a single line of diodes) and parallel scanned across the observer's viewing plane. Consider the system shown in Figure 11.5. As the scan traverses the image plane, a current is applied to each LED and the corresponding luminescent signals are traced out in lines. An eyepiece adjusts the LED image to a distance from the viewer that is comfortable.

The scanning occurs at a rate higher than the rate at which flicker is seen by the human eye (over 20 Hz). Most LED displays, as used in I^2R systems, form the back end of the EO mux subsystem (as described in Chapter 9). Here, the scan occurs at the same rate as that of the infrared linear-detector array scan. The vertical angular size of the image depends on the array height and the eyepiece focal length:

$$\alpha_y = d_y / f_e \tag{11.9}$$

The vertical image size is related to FOV_v by the magnification

$$\alpha_y = M \times FOV_v \tag{11.10}$$

In the system shown in Figure 11.5, a and b describe the size of the LEDs and c describes the distance between the centers of the LEDs in the vertical direction.

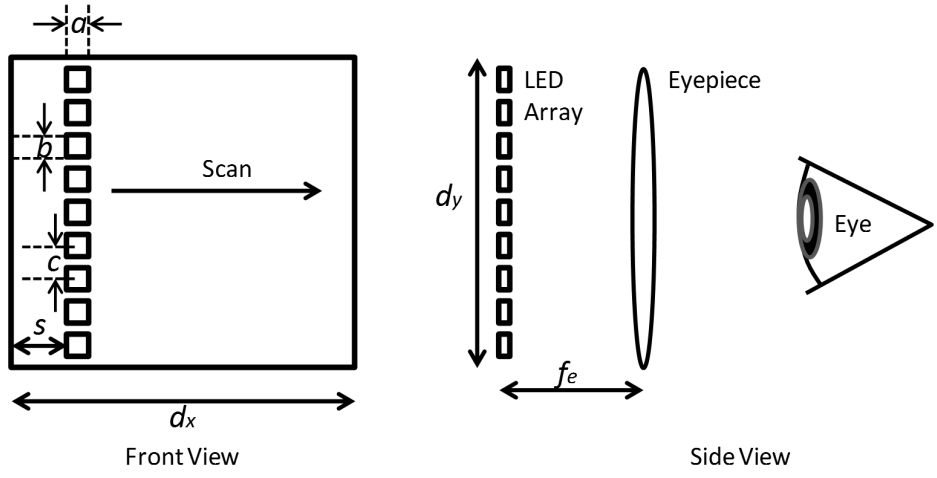


Figure 11.5 LED array display.

The distance of the LED array from the left side of the image plane is described as s. Many scanned LED displays are analog in nature (i.e., the voltage and corresponding luminescent signal are continuous across the scan), so there is no sampling in the horizontal direction. Therefore, the spatial function for the scanned array shown is

$$o(x,y) = \left[i(x,y) \frac{1}{c} comb\left(\frac{y}{c} \delta(x-s) \right) \right] ** rect\left(\frac{x}{a}, \frac{y}{b} \right) \tag{11.11}$$

Here, the spatial parameters x and y are in meters in the image plane and must be converted to angular object space. This is performed using the following equation:

$$o(x,y) = \left[i(x,y) \frac{d_y}{cFOV_v} comb\left(\frac{yd_y}{cFOV_v} \right) \delta\left(x - \frac{sd_x}{FOV_h} \right) \right] ** rect\left(\frac{xd_x}{aFOV_h}, \frac{yd_y}{bFOV_v} \right) \tag{11.12}$$

where x and y are in radians (object space). In addition, d_x and d_y are the width and height of the displayed image. The spatial frequency domain can now be compared directly with all of the other system contributors. The Fourier transform of (11.12) gives the spatial frequency representation of the display:

$$O(\xi,\eta) = \left[I(\xi,\eta) ** comb\left(\frac{cFoV_v \eta}{d_y} \right) \right] \frac{abFOV_v FOV_h}{d_x d_y} sinc\left(\frac{aFOV_h \xi}{d_x}, \frac{bFOV_v \eta}{d_y} \right) e^{-j2\pi \frac{sFOV_h}{d_x}} \tag{11.13}$$

If the input signal (the image to be displayed) is band limited such that all frequency components are less than half of $d_y/(cFOV_v)$, then no aliasing occurs and the MTF of the LED array can be written as

$$MTF_{LED}(\xi,\eta) = sinc\left(\frac{aFOV_h \xi}{d_x}, \frac{bFOV_v \eta}{d_y} \right) \tag{11.14}$$

If the signal is not band limited, then aliasing occurs on the output image and artifacts are seen. Well-designed systems use the optics and detector MTFs to band limit the imagery so aliasing is controlled. Further discussion of acceptable image aliasing for human consumption is described later in this chapter.

11.4 Liquid-Crystal Displays

The LCD [4] comprises a thin, clear layer of twisted nematic crystals placed between two conducting plates that are transparent. Some common LCDs use crossed polarizers as the two conducting plates. With an applied voltage, the LCD material (twisted nematic crystals) then turns the light polarization to pass through the crossed polarizer. The degree of polarization twist determines the shade of brightness that passes through the polarizer. The LCD panel is an aggregate of these cells that form an electronically addressable display. LCDs can be small, light, and inexpensive.

LCDs are passive devices that modulate passing light. Therefore, they work as well in bright-ambient light as they do in dim lighting. Their performance is strongly influenced, however, by the spatial distribution of lighting. Also, LCD displays are usually viewed directly by the observer. LCDs require very little power dissipation because the power is only required to provide dynamic scattering in the LCD fluid. LCDs are a passive component and are typically backlit by some uniform source. LCDs are not as fast as LEDs and typically require tens of milliseconds to turn off or on compared with some LEDs in the nanosecond realm. The slow response time limits their utility in high-speed applications.

Advantages of LCD displays are that they are compact and light, have low power consumption and negligible distortion, and do not flicker as long as there is a constant backlight. Disadvantages of LCDs are that have a limited viewing angle and smearing artifacts caused by slow response times.

The spatial display function and the transfer function can be derived from a simple model shown in Figure 11.6. The LCD array is a rectangular grid of LCD cells surrounded by an inactive area. The ratio of the LCD active cell area to the total array area is called the *fill factor*. The spatial sampling of the input image and the display of the image on the rectangular functions can be written as

$$o(x,y) = \left[i(x,y) \frac{1}{cd} comb\left(\frac{x}{d}, \frac{y}{c} \right) \right] ** rect\left(\frac{x}{a}, \frac{y}{b} \right) \tag{11.15}$$

Taking this function to angular object space gives

$$o(x,y) = \left[i(x,y) \frac{d_x}{dFOV_h} \frac{d_y}{cFOV_v} comb\left(\frac{xd_x}{dFOV_h}, \frac{yd_y}{cFOV_v} \right) \right] ** rect\left(\frac{xd_x}{aFOV_h}, \frac{yd_y}{bFOV_v} \right) \tag{11.16}$$

where x, y, and the FOVs are in milliradians. This spatial function can be converted to the spatial frequency domain by taking the Fourier transform:

$$O(\xi,\eta) = \left[I(\xi,\eta) ** comb\left(\frac{dFOV_h \xi}{d_x}, \frac{cFOv_v \eta}{d_y} \right) \right] \frac{abFOV_v FOV_h}{d_x d_y} sinc\left(\frac{aFOV_h \xi}{d_x}, \frac{bFOV_v \eta}{d_y} \right) \tag{11.17}$$

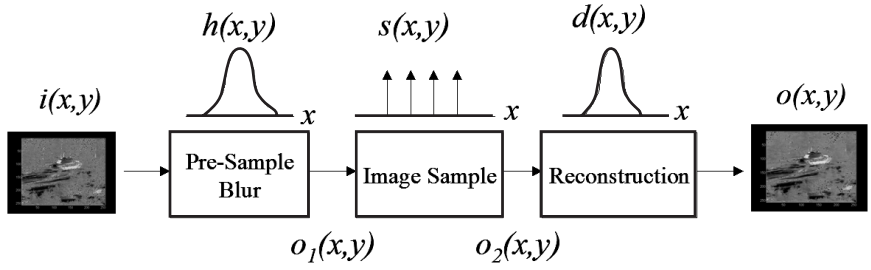


Figure 11.6 Sampled imaging system steps in space.

Note that this function is very similar to the spatial frequency representation of the LED array (11.13). The difference here is that the LCD array is in both the horizontal and vertical directions. Also note that if the image spectrum is not band limited in both directions, aliasing occurs *before* the display of the imagery. If the display is band limited to half of the *comb* function spacing, then the MTF is identical to that described in (11.14).

11.5 Plasma Displays

Whereas LCDs have been approved for military systems and have been integrated into weapons such as the M1 Abrams, plasma displays have been available for some time in the commercial market and are sometimes used in military applications (e.g., rapid prototyping systems). Each pixel in a cell is a mixture of noble gas and small amounts of mercury vapor between two panes of glass. A voltage is placed across the cell creating a plasma in a manner similar to fluorescent lamps. When electrons strike the mercury atoms, energy is converted to ultraviolet light that strikes a phosphorous surface that is coated on the glass wall of the cell. As with CRTs, different phosphors emit different colors of light in a red, green, and blue (RGB) fashion. The phosphors convert the ultraviolet light to visible and infrared light. Only around 40% of the emitted light from the phosphor is in the visible region and the infrared light is converted to heat in the glass. In a monochrome plasma display, the gas is usually neon that emits an orange color.

In the early 2000s, plasma displays were the largest commercial market for large flat-panel displays, but in roughly 2006, LCDs took over the market for the most popular large display. Advantages of plasma displays are superior contrast ratios since very dark blacks are achievable, wider viewing angles, and less motion blur due to fast response. Disadvantages are screen burn-in (image retention after long periods) and they require higher power and are less efficient than LCDs. In terms of blur and MTF, the plasma display can be treated in the same manner as an LCD display [see (11.15) and (11.17) with MTF equation, (11.14)].

11.6 Sampling and Display Processing

Displays are a primary component in the sampled image reconstruction process. Recall that some imagers do not sample the image (i.e., are digital), but in today's world most are sampled. Once the image is sampled at the detector, the sampled image is processed by electronics, shown on a display, and provided to the human eye. The reconstruction filter is comprised of electronics filtering, the display, and the eye. The display element smoothes or "filters" the digital signal and makes a continuous function for the eye to consumer. References [5–7] provide in-depth treatment of sampled imaging systems. From Chapter 3, the process can be described in Figure 11.6.

The source target and background are imaged through the atmosphere, the optics, and the detector. All of these components filter the image with a blur, so the component point spread functions are convolved with the input image. The

detector samples the image, so a *comb* function is multiplied by the blurred image. The reconstruction is a convolution of the reconstruction filter with the sampled signal (the weighted comb) to provide an output image.

In the frequency domain, the above operation is equivalent to an input spectrum multiplied by the presample MTF and that band-limited signal is convolved with a frequency domain *comb* function as shown in Figure 11.7. If the pitch of the detector sampling is in milliradians, the *comb* function in frequency provides a "replica" of the presampled spectra at one over the sample spacing in cycles per milliradian. In the figure, the replicas occur at 2 cycle/mrad, so the image was sampled at 0.5 mrad. The Nyquist rate is half of the sampling frequency of 2 cycle/mrad, so it is 1 cycle/mrad as shown. The FO+ and SO+ are the first-order and second-order replicas, respectively, caused by the frequency *comb*. The baseband signal corresponds to the original sampled spectrum. The display MTF (the case shown is a Gaussian MTF corresponding to a CRT display spot) is multiplied by the sampled spectrum to give the reconstructed spectrum shown in Figure 11.8.

The baseband is now the presampled blur spectrum multiplied by the display MTF. The replicas (FO, SO, and so on) are also multiplied by the display MTF, so the display is part of the reconstruction filter that is used to control these unwanted signals. Many scientists and engineers call this aliasing and it is manifested in two ways. The replica content below the Nyquist is in-band aliasing corresponding to classical aliasing in radio signals. These signals in an image are jagged edges, moved edges, warping, and other artifacts. The signals beyond the Nyquist rate (*out-of-band signals*) are display artifacts such as visible raster in a CRT and pixelization in LEDs, LCDs, and plasma cells. Note that an LCD, LED, or plasma MTF corresponding to a square display pixel provides a *sinc* MTF that allows through much more out-of-band aliasing. The spectrum shown in Figure 11.8 does not include the eye filtering that reduces the baseband and replica spectra even further. A full derivation of the sampled image transfer is provided in Chapter 3.

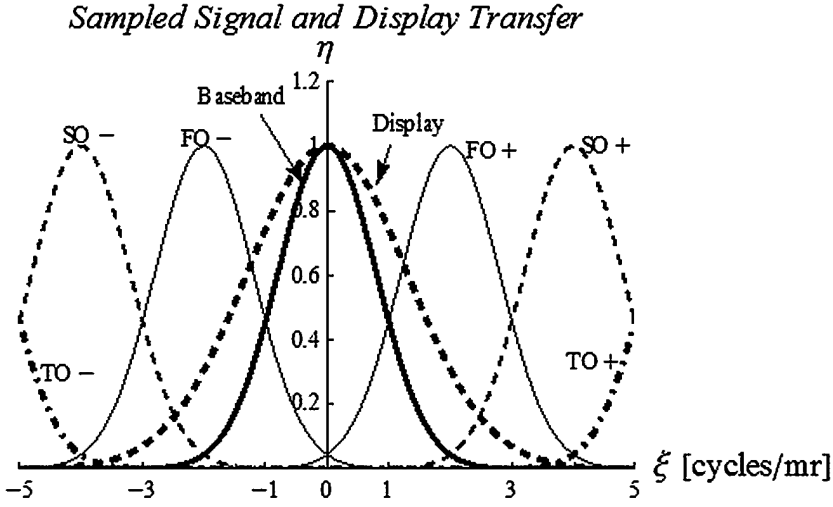


Figure 11.7 Sampled spectrum with display transfer.

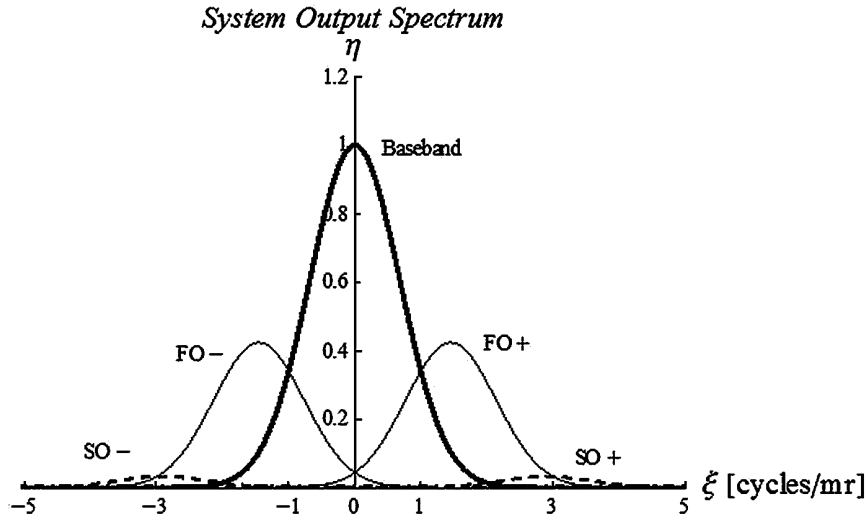


Figure 11.8 Reconstructed spectrum after display.

11.7 Human Perception and the Human Eye

The observer must interpret the information that is presented on a display. To more fully understand the response of the eye to this information, we describe the eye in physical terms and then provide models that can be used to estimate the eye MTF.

The human eye is similar to a camera. It forms a real image on a light-sensitive surface called the *retina* (see Figure 11.9). Light enters the eye through a transparent layer called the *cornea*, then traverses through an anterior liquid chamber called the *aqueous humor*. The light from this anterior chamber is limited by the *iris*, which is

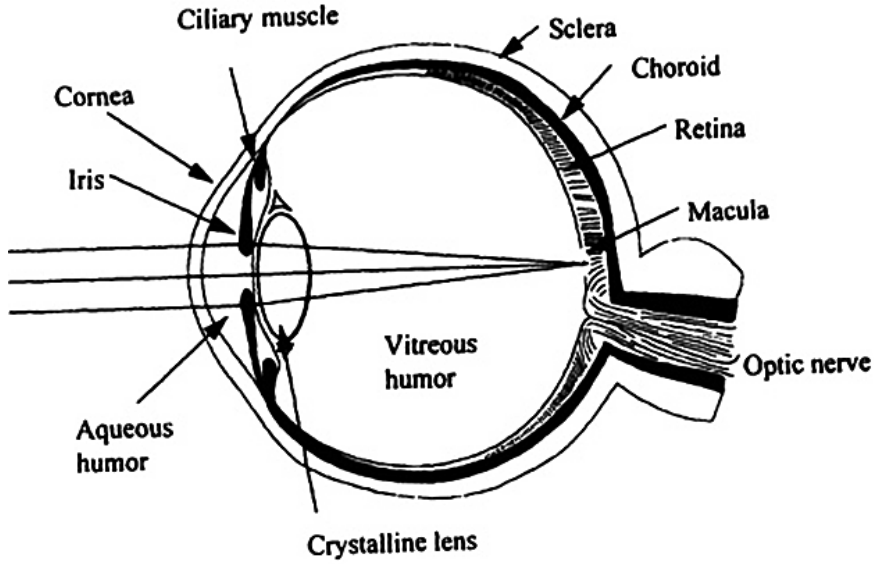


Figure 11.9 The human eye.

the effective entrance pupil of the eye. The diameter of the iris changes in response to the amount of light received by the eye. In reality, the pupil is not that important in the adaptation of the eye to large variations in light level. The most popular theory is that the eye adapts to large changes in light levels through chemical changes in the photoreceptors. Some of the light passes through the iris to a crystalline lens that is an elastic material. The lens power is determined by the ciliary muscles that pull on the lens radially to change the lens curvature. This change in optical power allows image focusing of near and far objects on the retina. After refraction by the lens, the light is imaged onto the retina through the vitreous humor.

The eye is a spherical jelly-like mass that is protected by a tough membrane called the *sclera*. The cornea is the only transparent part of the sclera; the rest is white and opaque. Inside the sclera is a dark pigmented layer called the *choroid*. The choroid is an absorber of stray light, similar to the black paint on the inside of EO devices. A thin layer of light-sensitive photoreceptors, like detectors, covers the inner surface of the choroid. The eye has two kinds of photoreceptors (125 million in all): *rods* and *cones*. Rods have the characteristics of high-speed, course-grain, black-and-white film and are extremely light sensitive. These photoreceptors can image light that is too dim for cones to process, but the images are not well defined. Cones are like low-speed, high-grain, color film that performs very well with bright light. The photoreceptor difference is one of the primary reasons that day and night vision provide two photometry curves: *photopic* and *scotopic* (see Chapter 5). At the retinal center, a depression in the photoreceptor layer is called the *macula*. It is around 2.5 to 3 mm in diameter with a rod-free region in the center of about 0.3 mm in diameter called the *fovea centralis*. Hecht and Zajac [8] state that the moon subtends a spot of around 0.3 mm on the retina. In this region, the cones are closely packed and provide the sharpest detailed images. The photoreceptors are ac coupled. The eye is continually moving the fovea centralis over objects of interest. The image is constantly shifted across the retina. If the image is held stationary on the retina, the image actually fades out.

The radii of curvatures [9] for the cornea, crystalline lens first surface, and crystalline lens second surface are 7.8, 10, and 6.0 mm, respectively. The indices of refraction of the aqueous humor, the lens, and the vitreous humor are 1.336, 1.413, and 1.336, respectively. An interesting note is that one of the reasons humans do not see well under water is that the water refractive index is typically 1.33, very close to the eye indices. Recall that little change in the refractive index at a curved interface provides little focusing power of the input light. The crystalline lens is fascinating in that it is a graded index lens. The index is 1.406 in the center, 1.375 at the equator, and 1.387 (aqueous) and 1.385 (vitreous) at the vertices. Focusing the lens is called *accommodation*, resulting in a change of the lens shape. The lens is almost spherical until pulled by the ciliary muscles, where the lens takes on a flatter shape. The front surface radius can change from 10 to 5.3 mm. The total power of the eye is around 60 diopters. The cornea–air interface accounts for around 42 diopters of the total power [10].

The age degradation of the crystalline lens and the ciliary muscles tends to affect eye performance in a significant way. The closest point that the eye can image is around 7 cm for a teenager, 25 cm for a young adult, and 100 cm for middle-aged people. Lenses are prescribed for people who cannot accommodate for the

necessary power. Eyeglass power is prescribed in *diopters* (recall that power is the reciprocal of the focal length). Myopia, or nearsightedness, is the case where the eye provides too much power, and a negative lens is prescribed to give a less overall system power. Hyperopia, or farsightedness, is the case where the eye does not provide sufficient power to image on the retina, so a positive power prescription is the solution. Ablation of the corneal surface with excimer laser radiation to change the surface curvature is becoming a common solution to either of these accommodation problems.

The band of sensitivity for the eye ranges from roughly 360 to 780 nm. There have been cases where people have seen images outside these wavelengths. The typical resolution of a human eye under certain conditions can be estimated by using the pupil size and a 0.5-μm wavelength with any one of the angular resolution criteria in the optics chapter. These estimates provide ballpark values, but are probably never realized because of imperfections in the eye. Typical angular resolutions of the eye range from 200 to 400 μrad.

11.8 Modulation Transfer Function of the Eye

The modulation transfer function of the eye is taken from Overington [11] and is the product of three distinct transfer function characteristics: *optical refraction*, *retina*, and *tremor*. These components are covered separately but must be cascaded (multiplied) to obtain the overall MTF of the eye.

The optical refraction MTF component corresponds to the limiting resolution of the pupil diameter. The pupil diameter, however, is a function of the display light level and can be estimated by

$$d_{pupil} = 4.30 - 0.645\ log(ftL) \quad [\text{mm}] \tag{11.18}$$

where ftL is the luminance of the display in foot-lamberts. There is 0.292 foot-lambert in 1 lumen/m^2-sr.

The optical MTF of the eye is estimated by

$$MTF_{eyeoptics}(\rho) = exp[-(43.69\rho/\rho_o)^{i_o}] \tag{11.19}$$

where ρ_o and i_o are transfer function variables that are related to the pupil diameter of the eye. Here, ρ is the circularly symmetric spatial frequency in cycles per milliradian. Table 11.1 gives the transfer function variable values for given pupil diameters. The pupil diameters given are for use with one eye. When two eyes are used, the pupil diameter becomes around 0.5 mm smaller. Also, the MTF given is in the eye angular space. The spatial frequency parameter ρ is increased [division of ρ in (11.19); recall the scaling theorem from Chapter 2] by a factor of the system magnification M in order to convert the MTF to the sensor object space.

The second MTF parameter associated with the human eye is the MTF of the retina. This parameter is similar to a detector MTF in that it is associated with the size of the retinal receptors and the focal length of the eye. A good approximation for the MTF is

Table 11.1 Eye Optical MTF Variables for Given Pupil Diameter

Pupil Diameter (mm)	ρ_o	i_o
1.5	36	0.9
2.0	39	0.8
2.4	35	0.8
3.0	32	0.77
3.8	25	0.75
4.9	15	0.72
5.8	11	0.69
6.6	8	0.66

$$MTF_{retina}(\rho) = exp[-0.375\rho^{1.21}] \tag{11.20}$$

where the spatial frequency must be converted to the angular object space with the magnification of the system in order to use the MTF in a system manner. Equation (11.20) gives the MTF of the eye in angular object space as seen by the eye. The sensor system angular object space is smaller by a factor of the system magnification M.

Finally, the last eye MTF component considered here is that of tremor, which can be estimated by

$$MTF_{tremor}(\rho) = exp[-0.4441\rho^2] \tag{11.21}$$

where the MTF has to be corrected for the system magnification for placement in sensor object space. In (11.19) through (11.21), ρ is replaced with ρ/M to give the MTFs in sensor, or object, space. Tremor is associated with the normal movement of the eye. Eye movements comprise high-frequency tremor, low-speed drift, and flick. All of these are small, involuntary movements that are necessary for vision. Experiments [10] have been performed where the eye was artificially stabilized with a contact lens projector. The conscious image appeared to fade out. Therefore, the eye can be considered an ac-coupled system.

Example 11.3

Determine and plot the eye's optical MTF using the Overington model. Assume that an observer is viewing a monitor of 10-fL brightness. The solution is given by first determining the pupil diameter, using (11.19):

$$d_{pupil} = 4.30 - 0.645\ log(10) \quad [\text{mm}] \tag{11.22}$$

or 3.655 mm. Using Table 11.1, the MTF variables are determined. Interpolating for a 3.655-mm pupil diameter, $\rho_o = 26.2$ and $i_o = 0.754$. The optical MTF associated with the pupil size is then estimated by (11.19) as

$$MTF_{eyeoptics}(\rho) = \exp[-(43.69\rho/26.2)^{0.754}] \tag{11.23}$$

This function is shown in Figure 11.10. Note that if two eyes were used, the pupil size would be smaller and the MTF would provide more degradation.

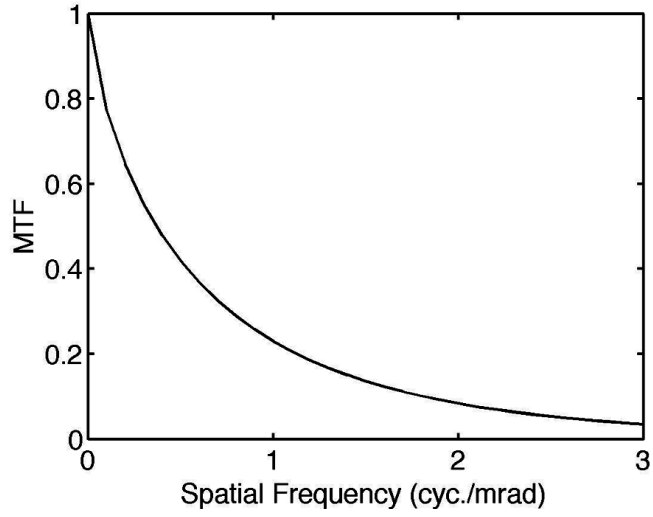


Figure 11.10 Example results.

11.9 Contrast Threshold Function of the Eye

In this second edition of *Introduction to Infrared and Electro-Optical Systems*, we depend more on the contrast threshold function (CTF) of the eye than we did in the previous version of the book. Therefore, we devote an entire chapter (Chapter 13) to the CTF and how it is translated to system terms. The CTF of the eye is a function that describes what spatial frequencies can be seen by the human and at what contrasts [12–15]. It is extremely important in the computation of sensor range performance when the human is an integral part of the system. The TTP, or target task performance, metric uses the system CTF (which includes the human eye CTF) to calculate the probabilities of detection, recognition, and identification as a function of sensor to target range.

11.10 Automatic Target Recognition

It has long been desired to automatically process the data provided by imaging sensors for a variety of functions. These functions range from such commercial applications as component inspection for manufacturing to military applications such as target detection and face recognition. The focus here is on military pattern recognition, though much of this material is applicable for both commercial and military systems.

The terminology that is often used to describe automated pattern recognition in the military community is generically referred to as *aided/automated target recognition* (Ai/ATR). Ai/ATR implies a sequence of mathematical operations where the result is target discrimination. The principal functions for the ATR sequence [16–19] are shown in Figure 11.11. The definitions for each of the ATR functions are provided in the following paragraphs.



Figure 11.11 ATR process.

Preprocessing is the conditioning of data. The input is typically an image such as an infrared image or a sequence of images, but it could be any sensor data. The preprocessing step conditions the data with processes like noise reduction/removal, image stabilization, image rotation, correction of dead pixels, and image registration.

Detection is a process whereby a portion of an image, usually called a *region of interest* (ROI), has been determined to contain an object of interest. Another definition is simply that an image successfully traverses through a prefilter (i.e., the detection process is a spatial screener). This process could be carried out by stepping a window through the entire image. The window is approximately the size of the target class of interest. At each step, a variety of tests (mathematical operations) are performed to determine if an object is present. Typical parameters of the window include contrast, various statistics, and power spectral density. Once a parameter is calculated, it is either compared with the background region around the window or compared with some predetermined threshold value. These values are determined based on some a priori knowledge of the target types of interest and by analysis of existing data that are similar to the test data. The primary purpose of detection is to eliminate most of the imagery data, as the processing power is usually not available to apply complex recognition algorithms to the large amount of imagery data. The real target decision is performed later.

Information extraction occurs after segmentation of the detection area. Once a potential target is localized through detection, it is extracted from the background as accurately as possible. Typical features that are extracted include the target's silhouette and internal features of the vehicle. Possible extraction includes spectral features, spatial features, temporal features, statistical features, and event alerts.

Classification is the process of assigning the potential target to a group. Detection and classification are the processes of discriminating between the target and the background. Recognition can be the process of determining whether a vehicle is a tank, truck, or armored personnel carrier. Identification can be the process of determining what type of tank a particular target is (e.g., a T-62, T-72, M1A1).

In evaluating an Ai/ATR algorithm, it is desirable to use measures that completely describe its performance. A wide range of factors must be considered including the environment in which the data were taken, the quality of the sensor data, the parameters of the sensor (resolution, sensitivity), and scene background clutter, to name just a few. A performance curve that is frequently used in the evaluation/design of Ai/ATR algorithms is shown in Figure 11.12. Historically, this curve has been known as the *receiver operating curve* (ROC).

As can be seen from the graph of Figure 11.12, the probability of target detection (this could be recognition or identification) is plotted with respect to the false alarm rate (FAR). A false alarm occurs when the Ai/ATR incorrectly classifies a nontarget as a target. As the probability of target detection (P_d) improves, the FAR

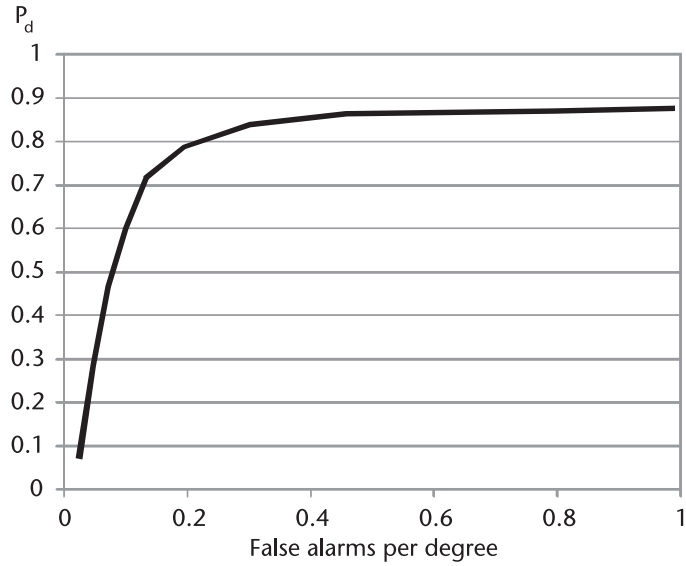


Figure 11.12 Ai/ATR performance evaluation curve.

increases. Ideally, one would like to have a very high probability of detection and a very low FAR. It is up to the algorithm designer to choose an acceptable P_d along with an allowable FAR.

In summary, tremendous improvements have been made in the supporting elements of ATR (sensors, processors, mathematics). Performance limitations can be attributed to the wide range of target signatures and backgrounds that vary tremendously depending on the test location and the environmental conditions. Despite its limitations, Ai/ATR is proving to be extremely useful in cueing an operator to an area of an image that contains an important target. This cueing and partial recognition are important for many military applications where timelines are critical and a large amount of imagery is involved. The commercial sector has had better success with optical character recognition (OCR) for typed pages and component inspection in manufacturing. The success in the commercial arena can largely be attributed to having a controlled environment.

11.11 Exercises

11.1 An imaging system has a vertical FOV of 2 deg and the imagery is displayed to a human observer. Assuming a typical viewing monitor angle, determine the system magnification.

11.2 A CRT display is a 525-line, 14-in. monitor that is given the shrinking raster test. The shrinking raster yields a total line height of 25 cm. Sketch the MTF of the CRT display in cycles per meter (or cycles per millimeter). Assuming that the monitor displays imagery from a 5-deg vertical FOV sensor, determine and sketch the CRT MTF in cycles per milliradian so that it can be compared with other system MTFs.

11.3 For Example 11.2, provide the necessary spot-size parameter for a cutoff frequency of 10 cycle/mrad given the same system.
11.4 Show that the cutoff frequency for the CRT is $1/\left(\sqrt{2\pi}\,\sigma\right)$.
11.5 Sketch the horizontal MTF for an LED display with an LED horizontal dimension of 0.05 mm, an LED scan distance of 2 cm, and a horizontal FOV of 5 deg. Sketch the MTF as a function of spatial frequency in cycles per milliradian.
11.6 Using the Overington model, determine the MTF of the eye that is viewing a display brightness of 20 fL. Sketch the results as a function of spatial frequency in cycles per milliradian. Compare these results with the MTFs of the retina and tremor.

References

[1] Pinson, L., *ElectroOptics*, New York City, NY: Wiley, 1985, p. 240.
[2] Waldman, G., and J. Wootton, *Electro-Optical System Performance Modeling,* Norwood, MA: Artech House, p. 170.
[3] Biberman, L., "Image Display Technology and Problems With Emphasis on Airborne Systems," in *The Infrared and Electro-Optical Systems Handbook*, Bellingham, WA: ERIM and SPIE Press, 1993.
[4] Goodman, L. A., "The Relative Merits of LEDs and LCDs, Proceedings for Society of Information Display," Vol. 16, No. 1, 1975.
[5] Vollmerhausen, R., "Impact of Display Modulation Transfer Function on the Quality of Sampled Imagery," *SPIE Proceedings of Aerospace/Defense Sensing and Controls*, Orlando, FL, 1996.
[6] Vollmerhausen, R., "Application of the Sampling Theorem to Solid State Cameras and Flat Panel Displays," *U.S. Army Night Vision and Electronic Sensors Directorate Report NV-2-14*, Ft. Belvoir, VA, 1989.
[7] Vollmerhausen, R. and Driggers, R., *Analysis of Sampled Imaging Systems*, Bellingham, WA: SPIE Press, 2000.
[8] Hecht, E., and A. Zajac, *Optics,* Reading, MA: Addison-Wesley, 1979, p. 138.
[9] Meyer-Arendt, J. R., *Introduction to Classical and Modern Optics*, Englewood Cliffs, NJ: Prentice-Hall, 1984.
[10] Yu, F. T., and I. C. Khoo, *Principles of Optical Engineering,* New York City,NY: Wiley, 1990.
[11] Overington, I., *Vision and Acquisition*, New York City, NY: Crane, Russak and Co., 1976.
[12] Blackwell, H., "Contrast Thresholds of the Human Eye," *Journal of the Optical Society of America*, Vol. 36, 1946.
[13] Karim, M., *Electro-Optical Displays*, NY: Marcel-Dekker, 1993.
[14] Holst, G., and A. Taylor, "What Eye Model Should We Use for MRT Testing," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing, SPIE*. Vol. 1309, 1990.
[15] Holst, G., *Electro-Optical Imaging System Performance,* Orlando, FL: JCD Publishing, 1995, p. 130.
[16] Haiken, S., *Neural Networks: A Comprehensive Foundation*, New York City, NY: Maxwell McMillan International, 1994.
[17] Keinosuke, F., *Introduction to Statistical Pattern Recognition, New York City, NY:* Academic Press, 1972.
[18] Madler, M., and E. Smith, *Pattern Recognition Engineering*, New York City, NY: Wiley, 1993.
[19] Ratches, J, "Review of current aided/automatic target acquisition technology for military target acquisition tasks," *Optical Engineering*, Vol. 50(7), 2011.

CHAPTER 12

Historical Performance Models

This chapter provides a review of past imaging system performance models. The models are not current, but it is worthwhile to understand the past models in that many scientists and engineers still use the nomenclature and even the calculations. A good deal of the material from this chapter comes from a review paper by Ratches, Vollmerhausen, and Driggers [1].

We cover here system models that describe the target acquisition performance of a human observer and an infrared imager. The models were adopted by the military infrared imaging community as an assessment of how well an ensemble of observers performs the tasks of target detection, recognition, and identification. The models were used in infrared imager design and assessment, where military users understood how the metrics predicted by the models related to system performance on the battlefield. This review begins with early work in the late 1950s and proceeds to other modeling successes (prior to the TTP metric). Understand that these models were the basis for many billions of dollars of sensor design to include soldier weapon sights, tank sights, and aircraft sensors.

12.1 Introduction

Infrared system performance model development has been ongoing for more than 50 years at a U.S. Army research and development laboratory. Although the U.S. Army leads these research efforts and maintains a current performance model, the work has been used by most infrared systems experts in the world. The U.S. Army effort resulted in a set of empirical analytic models that predicted the target acquisition performance of a system composed of a human observer and an EO imaging sensor. The emphasis of this chapter is on the subset of electro-optics known as thermal imaging. The resultant models have been adopted by the military infrared imaging community as quantitative assessments of how well an ensemble of observers, using infrared imagers, will perform the tasks of detecting, recognizing, and identifying tactical targets on a battlefield. The scientific investigation and engineering processes have produced models that are routinely used by Department of Defense agencies and their contractors to design and optimize military thermal imagers, frequently referred to as forward-looking infrared or FLIRs. Most importantly, the military users of these models understand how the metrics predicted by the models relate to how the systems will perform on a battlefield.

The effort began approximately 50 years ago at the then U.S. Army Night Vision Laboratory (NVL). Today NVL is known as the U.S. Army Communications & Electronics Command (CECOM) Night Vision & Electronic Sensors Directorate (NVESD) but it is still referred to as NVL in the community and that acronym will be used in this chapter. The models were a response to the army's need to develop

and field a whole family of night vision systems in order to fight under nighttime conditions. The modeling concepts that were investigated were driven by the outstanding FLIR system considerations at the various times during this 40-year span. Also, the thermal imaging sensor models are really only one type of model that has undergone a series of modifications or improvements that address more sophisticated aspects of thermal imaging systems. The initial model addressed simple detector arrays that scanned across the scene, but grew in complexity as detector arrays became focal plane arrays, resolution became nearly isotropic, system level noise grew more complex, sample data effects became more important, observer perception characteristics grew in importance, and other complexities arose that will be described in this chapter. A similar story can be told for models of other military EO systems, such as image intensifiers, albeit with their unique technical issues.

12.2 Johnson Model Fundamentals

The seminal hypothesis for the NVL performance models was proposed by John Johnson of NVL in 1958 [1]. Johnson's original concept was based on work done by Otto Schade [2] with televisions (the first man-made EO system). Johnson proposed that the ability of an observer to acquire military targets in scenes (detect, determine orientation, recognize, and identify) when viewing through an EO device is dependent on how well he can resolve bar patterns of varying frequencies through the device at the same contrast as the scene target-to-background contrast. Because thermal imagers did not exist when Johnson conducted his initial experiments, contrast was defined as target brightness (B_t) minus background brightness (B_b) divided by background brightness (B_b) for image intensifiers and television, or:

$$CONTRAST = (B_t - B_b)/B_b \tag{12.1}$$

When FLIRs became available, the concept was extended and contrast was replaced with target-to-background temperature difference (ΔT), which represented the signal input to thermal imagers:

$$SIGNAL = \Delta T = (T_t - T_b) \tag{12.2}$$

where T_b and T_t are the temperatures of the local (to the target) background and target, respectively.

A series of experiments was conducted with an ensemble of observers viewing through image intensifiers to determine how well they could resolve bar patterns and also perform the discrimination tasks of detection, target orientation, recognition, and identification of military targets. Detection is defined as discriminating the presence of an object of potential military interest from the background; orientation is the determination of the target aspect; recognition is determining the class of the target, for example, truck, personnel carrier, or tank; and identification is the determination of the member of the class, for example, M60, M48, or Stalin tanks. Johnson hypothesized that the ability of the observer–sensor ensemble to discriminate the targets was a function on how well a critical dimension on the target could be resolved. The critical

dimension for the various targets used was defined based on intuition but was usually chosen to be the minimum dimension. For combat vehicles this became the height of the vehicle. For a human target it was the width of a man. The discrimination level (e.g., detection, orientation, recognition, or identification) was then related to how many line pairs could be resolved across the defined critical dimension on the target for the limiting resolution measured with the observer–sensor combination. The number of line pairs resolvable (N) across a target critical dimension was calculated by multiplying the highest bar pattern frequency (f_b) that could be resolved at that contrast or ΔT by the observer–sensor ensemble times the target angular dimension (h). The angle h in milliradians was obtained by dividing the target critical dimension in meters by the target range in kilometers. The calculation of resolvable cycles is shown in (12.3). Table 12.1 shows the results of Johnson's experiments.

$$N[\text{cycles}] = f_b[\text{cycle/mrad}] \cdot h[\text{mrad}] \tag{12.3}$$

The original Johnson experiments were run at relatively high contrast targets and bar patterns. The performance measured corresponded to resolution-limited performance. The extrapolation to noise-limited conditions or performance under any contrast or temperature difference came with the definition minimum resolvable temperature difference (MRT). In 1969 Lloyd and Sendall [3] defined limiting resolution of bar patterns as the MRT and the measurement was standardized at NVL. MRT is the measurement of an observer's threshold in bar temperature difference above ambient for recognizing a four-bar pattern as a function of the bar frequency. That is, at a given bar frequency determined by the bar spacing [equal to 1/(2 × bar width)], the temperature difference of the bars above ambient was reduced until the observers could no longer distinguish four bars; the temperature difference was then increased until the bars were just visible. The methodology for getting from MRT to target acquisition performance is presented later in this chapter.

Table 12.1 Johnson's Data Relating Resolution in Lines Resolved Across the Target Critical Dimension to Discriminating Targets to Various Levels

Target	*Resolution (N) per Minimum Dimension*			
Broadside View	*Detection*	*Orientation*	*Recognition*	*Identification*
Truck	0.90	1.25	4.5	8.0
M-48 tank	0.75	1.2	3.5	7.0
Stalin tank	0.75	1.2	3.3	6.0
Centurion tank	0.75	1.2	3.5	6.0
Half-track	1.0	1.50	4.0	5.0
Jeep	1.2	1.50	4.5	5.5
Command car	1.2	1.5	4.3	5.5
Soldier (Standing)	1.5	1.8	3.8	8.0
105 Howitzer	1.0	1.5	4.8	6.0
Average	1.0 ± 0.25	1.4 ± 0.35	4.0 ± 0.8	6.4 ± 1.5

Much of the following modeling development has been published in the classified literature for infrared, the Infrared Information Symposium (IRIS). The original articles were largely classified due to the validation data sets, which indicated performance of new developmental infrared imagers. Most of the referenced articles are today declassified.

12.3 The MRT Model

The classical MRT model was first published by Ratches [4–6]. A recent historical perspective on the early model development appears in Ref. [7]. MRT is defined as the minimum temperature difference above 300K required by an observer viewing through the device to resolve a vertical four-bar pattern that has a 7:1 aspect ratio. The MRT is a monotonically increasing function of spatial bar target frequency, f_B, in cycles per milliradian. The expression for MRT is derived from the threshold signal-to-noise ratio required by an observer to resolve the pattern and is given by

$$MRT(f_B) = \left(\frac{\pi^2}{4\sqrt{14}}\right)\cdot\left(\frac{S}{N}\right)\cdot\left[\frac{NET}{H_{TOT}(f_B)}\right]\cdot\left[\frac{Vf_BQ(f_B)\Delta Y}{\Delta f_n F_R \eta_{OV} t_E}\right]^{1/2} \tag{12.4}$$

where *S/N* is the unitless threshold signal-to-noise ratio required to recognize one bar. Also known as SNRT, the value of *S/N* was found to be equal to 2.25 based on empirical results. NET is the noise equivalent temperature difference in kelvin and is calculated as follows:

$$NET = \frac{4F^2\sqrt{\Delta f_n}}{\pi\sqrt{A_d N}\,\tau_o\int_{\lambda_1}^{\lambda_2} D^*(\lambda)\dfrac{\partial L(\lambda)}{\partial T}\,d\lambda} \tag{12.5}$$

where

F = f-number of the optics (unitless)
Δf_n = noise bandwidth (Hz)
A_d = detector area (cm^2)
τ_o = average optical transmission (unitless)
N = number of detectors scanned and summed in series
$D^*(\lambda)$ = specific detectivity (cm-$\sqrt{Hz}$/W or Jones)
$\dfrac{\partial L(\lambda)}{\partial T}$ = partial of radiance (Planck's equation) with respect to temperature (W/cm^2-sr-μm-K).

The noise bandwidth is defined as

$$\Delta f_n = \int_0^\infty S(f)H_{elec}(f)H_{Meas}(f)\,df \tag{12.6}$$

where

$S(f)$ = normalized noise power spectrum (unitless)
$H_{elec}(f)$= electronic MTF (unitless)
and $H_{Meas}(f)$ = measurement device MTF (unitless).

At the time of the Ratches model, a first-generation FLIR (serial scanned) was the primary sensor for modeling and testing, so the measurement device was well defined with a particular MTF:

$$H_{Meas}(f) = 1/\left[1 + (f/f_o)^2\right]^{1/2} \tag{12.7}$$

and f_owas set to $1/2\tau_d$ and τ_d was the dwell time of the detector, which was defined as

$$\frac{1}{\tau_d} = \frac{\alpha\beta\eta_{OV}F_R}{\Delta X n \Delta Y \eta_{SC}} \tag{12.8}$$

where

α = horizontal field of view (FOV) (mrad)
β = vertical FOV (mrad)
η_{ov} = overscan ratio [ratio of the vertical instantaneous field of view (IFOV) to the vertical sample spacing] (unitless)
F_R = frame rate (sec^{-1})
ΔX = horizontal IFOV (mrad) found by dividing the detector horizontal dimension in meters by the focal length in meters × 1000
ΔY = vertical IFOV (mrad)
n = number of detectors in parallel (unitless)
η_{SC} = scan efficiency (unitless).

Returning to (12.4), V is the scan velocity of the scan mirror (mrad/s), t_E is the integration time of the eye (approximately 0.2 sec), and finally, $Q(f_B)$ is the spatial integration of the eye over a bar:

$$Q(f_b) = \int_0^\infty S(f_x)H_N^2(f_x)H_{WB}^2(f_x)H_{EYE}^2(f_x)\,df_x \tag{12.9}$$

The MTF terms in the eye integral are:

$H_N^2(f_x)$ = MTF of all system components after the detector including the display
$H_{EYE}^2(f_x)$ = eye MTF
and $H_{WB}^2(f_x)$ = Fourier transform of the bar width.

Since the bar is a *rect* function in space, then

$$H_{WB}^2(f_x) = sin(\pi f_x W_B)/(\pi f_x W_B) \tag{12.10}$$

where W_B is the bar width and is given by $1/(2f_B)$.

The MRT equation can be simplified with substitution of the NET of (12.5) into (12.4):

$$MRT(f_B) = \left[\frac{2.25\pi F^2}{\sqrt{14}\tau_o H_{TOT}(f_B)\int_{\lambda_1}^{\lambda_2} D^*(\lambda)\dfrac{\partial L(\lambda)}{\partial T}\,d\lambda}\right]\cdot\left[\frac{Vf_B Q(f_B)\Delta Y}{F_R\eta_{OV}t_E A_d N}\right]^{1/2} \tag{12.11}$$

A final reduction can occur since the scan velocity is the horizontal FOV times the frame rate

$$V = \alpha \cdot F_R(\text{mrad/s}) \tag{12.12}$$

so that

$$MRT(f_B) = \left[\frac{2.25\pi F^2}{\sqrt{14}\tau_o H_{TOT}(f_B)\int_{\lambda_1}^{\lambda_2} D^*(\lambda)\dfrac{\partial L(\lambda)}{\partial T}\,d\lambda}\right]\cdot\left[\frac{\alpha f_B Q(f_B)\Delta Y}{\eta_{OV}t_E A_d N}\right]^{1/2} \tag{12.13}$$

Note that the performance is driven by the eye integration time. A low frame rate is accompanied by a small noise bandwidth, a high dwell time, and the eye integrates only a small number of frames. A high frame rate is accompanied by a large noise bandwidth, a small dwell time, and the eye integrates a large number of frames. These conditions are equivalent in the human detection of bar targets.

12.4 The First FLIRS and Models

The MRT related system design parameters, such as D^*, optics diameter, and detector size, to subjective observer recognition of four-bar patterns. The Johnson model could then be applied to MRT for a given target signature, atmospheric condition, and discrimination level to predict field performance. This became critically important during the late 1960s and 1970s when the U.S. Army was ready to design, engineer, produce, and field a whole family of thermal night sights across all weapons systems. A performance model was needed that could relate FLIR design parameters to field performance by the soldier using the device. To optimize the design and choose the best contractor candidate, a method was needed to relate, quantitatively, system parameters such as detector sensitivity, detector MTF, optical *F*/#, optical MTF, electronics MTF, and display MTF to how well a soldier could acquire targets.

At this time a team of modelers at NVL led by Dr. Walter Lawson was attempting to build on the work of Johnson to come up with a performance model for FLIR systems. The objective was to generate an engineering model that predicted the probability of an ensemble of observers being able to detect, classify (tracked vs. wheeled vehicle), recognize, and identify tactical targets as a function of range, environment, and system parameters. With the derivation of the MRT expression [(12.4) and (12.13)], it became straightforward to apply the Johnson concept and

relate system parameters to performance in the field. The MRT provided the connection between Johnson's concept of resolvable bars across the target critical dimension and a system-level measurement and it was a measurement that could be routinely carried out in the laboratory. The MRT measurement also included the subjective performance of the observer. (A minimum detectable temperature difference (MDT) was also defined that would be the connection between the SNR of a "blob" and the ability to detect hot spots without any higher level discrimination. A measurement was performed in which the threshold of an observer was measured for detecting the hot target as a function of target size [8].)

The target acquisition performance model [9] that used the Johnson concept and MRT is shown in Figure 12.1. An inherent target signature in terms of a temperature difference (ΔT) of the target above the local background is attenuated by the atmospheric propagation to give an apparent temperature difference ($\Delta T'$) at the sensor. This $\Delta T'$ corresponds to a bar pattern frequency f', which the observer–sensor system can resolve through the MRT curve. The number of resolvable bar pattern cycles from (12.3), which can be resolved across the target critical dimension h (meters) at range R (kilometers), is then given by

$$f' \frac{h}{R} = N \tag{12.14}$$

This number of resolvable cycles N across the target critical dimension could then be related to probability of any level of discrimination through a set of empirically generated curves. These target transfer probability functions (TTPFs) were generated from an extensive set of field exercises given in Ref. [7]. The TTPFs represent the percentage of the ensemble of observers who could correctly perform the discrimination task, for example, detection, classification, recognition, and

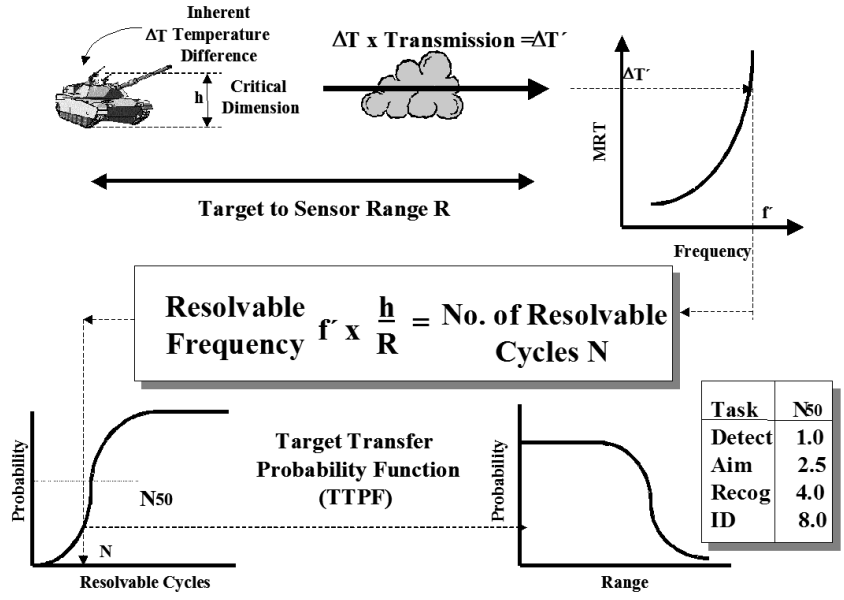


Figure 12.1 Model to predict performance based on Johnson and using system MRT.

identification. The question of false alarm rate is frequently brought up with respect to this performance. However, these experiments were designed such that the FAR was extremely low. The observers were instructed to respond only when they were very confident of their response. The number N was a function of target signature (ΔT and h), atmospheric propagation, and system parameters through the MRT.

The target ΔT was calculated from a measured or predicted signature by calculating the area weighted ΔT for the whole target from the ΔT's and areas of each component of a target signature when broken up into subareas of constant temperature. The critical dimension of the target was usually the height; however, that was not always the case. For example, for a man target, the width was used. The choice of critical target dimension for new targets is usually left to a panel of experts with a great deal of experience in the performance of FLIRs. Tables of critical dimensions have been published in the literature on the NVL model. Atmospheric propagation is routinely calculated using some standard model, such as LOWTRAN [10] (or the newer version, MODTRAN). The TTPF curves were generated from field performance data and one standard free-hand fit was used to match all levels of discrimination. The one curve was translated horizontally over resolvable cycles and the position of the curve was specified by N_{50}, which represents a 50% probability. The N_{50} then specified the entire curve. For example, the N_{50} values for detection, aiming (e.g., a missile gunner being able to put a cross-hairs on a target with sufficient accuracy to fire a missile), recognition, and identification were determined to be 1.0, 2.5, 4.0, and 8.0, respectively during this time frame. Later, as more field data were acquired, the values evolved.

The entirely new model that had to be developed in order to implement this modeling approach described in Figure 12.1, which related MRT to field performance, was called the ACQUIRE [9] model. The ACQUIRE model related FLIR system design parameters through the MRT to observer–sensor performance in the field. These included detector noise, optical transfer functions, detector transfer functions, electronics parameters, display characteristics, and observer eye integration time.

Other observer factors influence target acquisition performance, such as training, motivation, and reward. These factors have never been incorporated into the model. ACQUIRE could be used as a sensor design and contractor sensor selection tool. It was accepted by the community because it was validated with real field data, was well understood, and was promulgated throughout the national and international infrared community. Most importantly, as the data from laboratory-measured MRT and field measurement validation grew, the army decision makers could rely on the fact that a given thermal imager set of design parameters resulted in a required level of target acquisition performance on the battlefield. Costly field test verification of military-required performance was, thus, avoided and replaced with a relatively simple test that could be done on the production line.

The selection of the appropriate N_{50} to a specific situation was not always straightforward. Oftentimes an "expert" was required to select the N_{50} for a particular task and the target critical dimension. In addition, all imagers at this time were linearly scanned in the horizontal direction and a vertical MRT was not defined due to the sampling effects. Hence, the model ignored resolution in the vertical direction. Notwithstanding, the MRT model was used as the basis for field performance

predictions with useful fidelity. The MRT model was shown to give predictions that were, generally, representative of what was measured in the laboratory for linearly scanned thermal imagers. Typically, the predicted MRT curve crossed the measured system curve at some intermediate bar frequency and was optimistic or pessimistic at low frequency and the opposite at high frequency. Field performance was shown to be within ±20% in range for a given probability. To extend the MRT model to staring systems, an arbitrary cutoff of the MRT at one-half the theoretical limiting frequency (one over the detector IFOV) was imposed. This was done in order to account for the fact that a staring imager could not resolve four bars beyond that frequency, although there was modulation in that region of frequencies. This shortcoming of the MRT model is addressed in the next sections.

12.5 Model Improvements for Resolution and Noise

By the mid-1980s a new generation of thermal imaging was being developed. Linearly scanned 2D arrays of detectors and staring arrays were becoming available. Material and growth improvements coupled with the increases in sensitivity to be realized from time-delay-and-integrate (TDI) and staring provided the opportunity for a quantum leap forward in performance for thermal imaging. With the advent of second-generation scanning systems and staring sensors, shortcomings in the FLIR performance models became critical. System-level noise became as important as the detector noise. Noise introduced by detector-to-detector nonuniformity, the scanning/framing processing, multiplexing, fixed patterns, and electronic processing had to be considered in performance models for second-generation devices if the performance was to be accurately assessed. The 2D scanned and staring arrays were sampled in two directions with electronic multiplexing and the vertical resolution approached horizontal.

The other major improvement in second-generation systems over first-generation systems was in digitization. The detector array became a focal plane array of IR sensitive detectors with readout circuits bonded to the detector array and

Table 12.2 Temporal and Spatial Noise in Second-Generation Thermal Systems

3-D Noise Component Description		
Noise	*Description*	*Potential Source*
σ_{tvh}	Random spatiotemporal Noise	Basic detector temporal noise
σ_{tv}	Temporal row noise	Line processing, 1/*f*, readout
σ_{th}	Temporal column noise	Scan effects
σ_{vh}	Temporal spatial noise	Pixel rocessing, Detector-to-detector Non-uniformity
σ_{v}	Fixed-row noise	Detector-to-detector nonuniformity, 1/*f*
σ_{h}	Fixed-column noise	Scan effects, detector-to-detector nonuniformity
σ_{t}	Frame-to-frame Noise	Frame processing
S	Mean of all components	

which multiplexed the signal out of the dewar. This signal could now be digitized with analog-to-digital (A/D) conversion and processed using state-of-the-art digital processing. The major military interest in digital processing was to implement Ai/ATR in the sensor package. Second-generation systems scanned focal plane arrays that are going into the army's Horizontal Technology Integration FLIR B-Kit were designed to facilitate ATR on weapons platforms. The B-Kit was a set of infrared system components that was intended to replace first-generation infrared systems in a vehicle-independent manner. FLIR B-Kits have nearly isotropic resolution in both dimensions, no interlace, an improved SNR through TDI, and can be sampled at greater than once per detector dwell time in order to provide processors with the most computer friendly image possible in order to perform automated functions.

The NVL modeling group was led at this time first by Luanne Obert and then by John D'Agostino and addressed the noise and increased resolution issues directly [11, 12]. The group also implemented improvements in MRT modeling with the model FLIR90 and subsequently FLIR92 [13]. D'Agostino hypothesized that the total system noise could be reduced to eight components depending on whether the displayed noise had temporal variation t or spatial horizontal h or vertical v variation in the plane of the display. The standard deviation of each noise component σ represented a real displayed and measurable noise. Table 12.2 shows the eight components and their description and potential source for that noise component. The development of digital processing permitted the measurement of each of these components in a system in a laboratory. The new noise concept was implemented mathematically in the FLIR92 MRT equation. The validity of the 3D noise model for temporally coherent, spatially random noise was recently demonstrated [14].

Assuming statistical independence of the noise components, the total system noise in this 3D formulation for noise as a function of frequency f_s is the square root of

$$\Omega(f_s) = \sigma_{tvh}^2 E_t E_{vz} E_{hz}(f_s) + \sigma_{vh}^2 E_{vz}(f_s) E_{hz}(f_s) + \sigma_{th}^2 E_t E_{hz}(f_s) + \sigma_{tv}^2 E_t E_{vz}(f_s) + \sigma_v^2 E_{vz}(f_s) + \sigma_h^2 E_{hz}(f_s) \tag{12.15}$$

where E_t, E_{hz}, and E_{vz} represent the eye/brain temporal and spatial integration associated with each noise component and the z subscript indicates vertical or horizontal orientation of the four bars. The temporal and spatial integrators are approximately given by

$$E_t = \alpha_t / (F_R t_E) \tag{12.16a}$$

$$E_{hz} = \alpha_h / \left[R_h L_{hz}(f_s) \right] \tag{12.16b}$$

$$E_{vz} = \alpha_v / \left[R_v L_{vz}(f_s) \right] \tag{12.16c}$$

where R_h and R_v are horizontal and vertical sampling rates (sample/mrad); $L_{hz}(f_s)$ and $L_{vz}(f_s)$ are spatial integration limits (mrad^{-1}), which are approximately the hori-

zontal and vertical dimensions of the bar target; and α_h, α_v, and α_t are the sample correlation factors, which are equal to 1 for staring systems and may be greater than 1 for scanning systems; F_R is the frame rate; and t_E is the eye integration time.

The term σ_{tvh} is the basic detector noise normally characterized by the NET. It is related to the actual system bandwidth and not the artificial standard bandwidth used to measure NET. It becomes the NET when multiplied by the ratio of the equivalent noise bandwidth divided by the actual noise bandwidth.

A second innovation that D'Agostino introduced into the NVESD FLIR and target acquisition models was the use of resolution in both horizontal and vertical directions. Background experiments were performed at NVESD using simulated imagery. The experiments showed that more accurate performance predictions were made when resolution in both image directions was included. To preserve the existing well-understood Johnson concept for imaging and to have as little as possible impact on the well-established approach to predicting field performance from MRT, a 2D MRT was defined that does not correspond to a physical measurement that could be performed on a FLIR. A fictitious MRT function was defined whose temperature difference value was defined as that value at a frequency equal to the square root of the product of the horizontal and vertical MRT frequencies for the ΔT value measured on the horizontal and vertical MRTs. The prediction of field performance was then identical to that shown in Figure 12.1; however, the critical dimension of the target now became the square root of the target area in order to have a consistent 2D approach. The target area is the projected area of the target on the display. This means that the area can be different in different spectral regions due to the fact that different components of a target show up differently in the various spectral regions. The canvas of a 2.5-ton truck may be at the ambient temperature and have no temperature difference in the infrared. It would show up in the visible region. Figure 12.2 shows this 2D model diagrammatically. Although, conceptually, one might envision this approach as using "resolvable pixels" on the

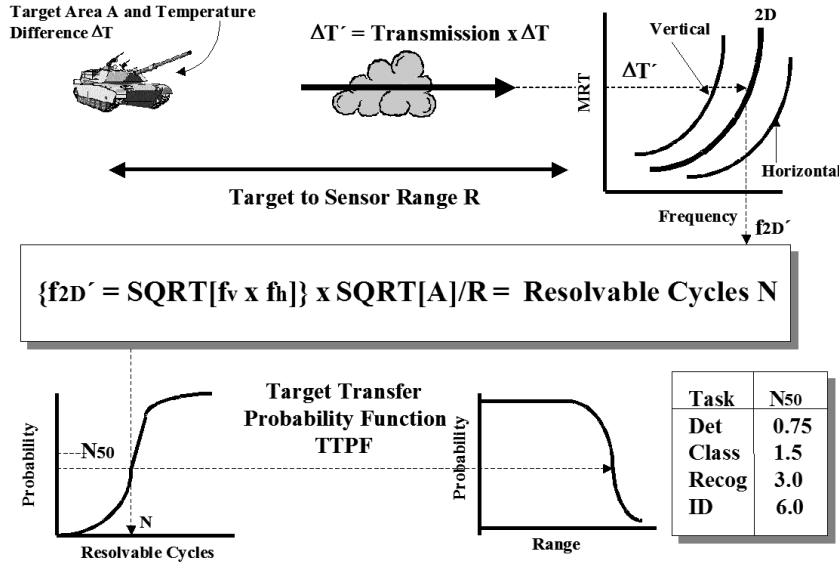


Figure 12.2 Two-dimensional approach to performance prediction.

target, the fundamental metrics are still linear 1D frequencies (square root of horizontal and vertical frequencies) and length (square root of area).

It is important to note that at this time when 2D resolution was being introduced into the model, more system component transfer options were introduced into the model also. Transfer functions for digital filters, CCD charge transfer, display sample and hold, and other electronic filters were added to the potential possible characteristics of model descriptions for candidate thermal imagers.

The FLIR MRT model with 3D noise and 2D MRT was released to the community under the name of FLIR92. The calculation of target signature, atmospheric propagation, and field performance has been released under the name ACQUIRE [9]. FLIR92 also includes the calculation of MDT, and ACQUIRE uses the MDT to compute "hot spot" detection ranges. ACQUIRE contains tables for the area, critical dimension for tactical targets, and an analytic curve fit to the TTPFs. The TTPF N_{50} values had to be changed somewhat in order to validate the model for second-generation systems and to revalidate the 2D model to the old database of performance with first-generation FLIRs. The original NVESD model values for N_{50} are shown in Table 12.3 compared to the new values used in ACQUIRE. Also, the "aim" discrimination level is dropped and a classification level is introduced. The new values for N_{50} brought closer agreement with the values used in image intensifier modeling. This was aesthetically pleasing since it brought the modeling of different EO technologies into closer agreement. Image intensifiers have isotropic resolution in all directions and the recognition criterion for them has been $N_{50} = 3$ for many years, and now FLIR models had the same criterion because their resolution became isotropic in two dimensions.

It is important to note that FLIR92 did not account for sample data effects any differently than the original model. An arbitrary asymptote was imposed on a staring system MRT at the Nyquist frequency. This is important in the modeling activity discussed later.

Two complementary developments to the modeling helped enable the significant improvements in the capability of the FLIR 92 and ACQUIRE models. These were the development of an Advanced Sensor Evaluation Facility [15] and the expansion of the applicability and use of perception testing [16, 17]. Laboratory testing of thermal imaging systems was upgraded in order to make use of digital processing. Signal trains in an imaging system could by digitized and processed in order to help characterize the system. Use of such equipment as frame grabbers enabled the isolation of the various noise components, which permitted the validation of the noise modeling concepts. In addition, the ability to generate large amounts of simulated

Table 12.3 Changes in Johnson Criteria from Original NVESD Model to ACQUIRE

Resolvable Cycles across Target Critical Dimension					
Task	*Detect*	*Aim*	*Classify*	*Recognize*	*Identify*
Original N_{50}	1.0	2.5	—	4.0	8.0
ACQUIRE N_{50}	0.75	—	1.5	3.0	6.0
N_{50} for a man target = 0.75					

From [18].

targets that could be used as input stimuli to automated perceptual testing enabled the generation of large amounts of target acquisition data. Many independent and dependent variables to observer-in-the-loop performance under controlled conditions could be studied that provided new in-depth understanding and new concepts for advanced models. The results from the perception testing had high statistical significance due to the large number of replications that could be performed under the controlled environments.

12.6 Incorporating Eye Contrast Limitations

The FLIR92 model provided accurate performance predictions for first- and second-generation thermal imagers. However, the early 2000s saw significant advances in the development of sensitive staring detector arrays, and these arrays were being incorporated into a wide variety of sensor systems. Staring arrays had characteristics that led to errors in FLIR92 performance predictions. Due to the sensitivity of the new staring arrays, the contrast limitations of the eye were important for establishing performance limitations, and these were not modeled in FLIR92. The addition of eye contrast limitations to the model is described below. Also, the performance loss due to sampling artifacts was typically not as severe as that modeled by FLIR92. The change in the sampling theory that was also applied is described in this chapter.

An upgrade to the thermal model, called NVTherm, incorporated two changes in the eye model. The theory behind these changes is described in Refs. [19] and [20], and the two changes are briefly described here.

The first change involved replacing the fixed signal-to-noise threshold at which a bar pattern is assumed to be detected with a variable threshold that varies with display luminance and the spatial frequency presented to the eye. In the theory described by (12.13), the signal-to-noise ratio threshold (SNRT = 2.25) at which a bar pattern is detected is assumed to be constant regardless of display luminance or spatial frequency. The experiments of Rosell and Wilson in the early 1970s, and experience with image intensifier sensors at NVL during the same period, demonstrated that SNRT was not fixed [21]. NVTherm incorporated a variable SNRT that was based on the measured contrast threshold function (CTF) of the eye and visual system.

The CTF is one of the most common ways of characterizing human vision. The following procedure was used to measure CTF. The observer viewed a barchart or sine wave pattern. While holding average luminance to the eye constant, the contrast of the bar pattern was lowered until the pattern was no longer visible to the observer. That is, the dark bars were lightened and the light bars darkened, holding the average constant, until the bar–space–bar pattern disappeared. The contrast was then increased until the bars were once again visible. The average contrast between where the bars disappeared and reappeared was defined as threshold contrast for that bar pattern at that adapting luminance. The procedure was repeated for various bar spacings—that is, for various spatial frequencies. The function of threshold contrast versus spatial frequency at each light level was called the CTF at that light level.

When modeling a system, the display luminance and display size are specified. For the specified display luminance, the CTF measured for that luminance is an indicator of the eye's ability to integrate signal and discriminate noise at each spatial frequency.

The fixed SNRT in (12.13) was replaced with the following variable SNRT:

$$SNRT \Rightarrow K_{eye} CTF(f) \tag{12.17}$$

where K_{eye} was a constant. Eye MTF and variations in eye integration time due to light level variations were included in the CTF, and these factors no longer explicitly appeared in the modified MRT formula shown in (12.18).

Eye contrast limitations affected the MRT equation in a second way. FLIR92 and previous models predict MRT based on the sensor noise and MTF characteristics. Kornfeld and Lawson recognized that, in the limit as sensor noise decreases, the eye's noise or contrast threshold limitations would be the actual limit on performance [22]. In NVTherm, a second term was added to the MRT equation such that performance was limited by the eye as sensor noise becomes small.

With both of these changes incorporated, (12.13) for sensor system MRT became:

$$MRT(f_B) = \left[K_{eye}^2 CTF^2(f_B) \beta^2(f_B) + \frac{4 S_{temp}^2 CTF^2(f_B)}{H_{TOT}(f_B)} \right]^{1/2} \tag{12.18}$$

where:

$$\beta(f_B) = \left[\frac{\pi F^2}{\sqrt{14} \tau_o H_{TOT}(f_B) \int_{\lambda_1}^{\lambda_2} D^*(\lambda) \dfrac{\partial L(\lambda)}{\partial T} \, d\lambda} \right] \cdot \left[\frac{\alpha f_B Q(f_B) \Delta Y}{\eta_{OV} A_d N} \right]^{1/2}$$

and

$\beta(f_B)$ = MRT(f_B) given in (12.13) without the 2.25 (SNRT) or t_E
$CTF(f_B)$ = the contrast threshold function of the eye based on measured data
K_{eye} = eye threshold calibration constant, which replaces SNRT; its value is 1050, which seems large, but the CTF of the eye is about 0.002 at peak sensitivity
S_{temp} = the scene thermal contrast, which results in the average display luminance; typically, S_{temp} was equal to half of the sensor dynamic range.

The MRT equation now consisted of two terms. The first term represented the contribution of sensor noise; the second term represented the contribution of "eye noise" or contrast threshold limitations. The relationship between these two terms depended on the amount of noise generated by the infrared detectors and on the setting of the sensor gain control. As the average scene temperature heated up, the sensor gain was reduced and the eye contrast term began to dominate. As the scene cooled off and gain was increased, for intrinsically noisy detector arrays, the sensor noise dominated. In the limit of low detector noise or high scene thermal contrast, the MRT was dominated by the sensor optical blur and eye contrast limitations. In the limit of high sensor noise or very low thermal contrast, the sensor noise dominated the MRT.

The addition of eye contrast limitations to the MRT equation resulted in more realistic performance predictions from ACQUIRE. For example, to maintain dark adaptation, display luminance was normally kept low during night operations. The new model correctly predicted performance under these realistic conditions. Further, the performance of new, sensitive staring arrays was then modeled correctly, because the contrast limitations became more important as the detector noise decreased.

12.7 Model Improvement to Add Sampling

In the mid-2000s, the mathematical characterization of image artifacts that resulted from sampling, and the target acquisition performance loss that resulted from those artifacts, was described in several published books and papers [23–28].

Mathematically, the sensor and display were characterized by a sampled imager response function. The response function for a sampled imager was derived by examining the image formed on the display by a point source of light in the scene. The response function provided a quantitative way to characterize both the quality of the sampled imager's transfer response and its tendency to generate sampling artifacts.

The sampled imager response function depended on the sensor pre-sample MTF, the sample spacing, and the post-sample and display MTF. These sensor characteristics were known to the design engineer or system analyst. The sampled imager response function did not depend on the image samples, but rather on the process by which the samples were taken and displayed.

Since the sampling artifacts produced by an imager depended on the scene being imaged, one might question a mathematical process that quantified sampling artifacts without including an explicit description of the scene. In that regard, assumptions identical to those used for non-sampled imagers were considered.

The MTF was widely used to characterize a non-sampled imager. The MTF was also the Fourier transform of the displayed point spread function. It described the blur produced in the image by a point in the scene. In actual usage, the importance of a good MTF response at high frequency could not be established until the high-frequency content of the scene had been established. The real impact or importance of the sensor blur was not known until the scene content and task were known. Nonetheless, the MTF was proven to be a good indicator of the overall utility of the sensor. A high-frequency response in a non-sampled imager was a prized characteristic because of the possibilities it provided, not because a good MTF was always important in accomplishing every task when looking at any scene. Experience had shown that the MTF was a good way to characterize the quality of an imaging system. An image could not be defined until the scene had been described, but the characterization of the imager's response to a point source provided a good indication of the quality of images that could be expected under a variety of environments.

A similar logic applied to sampled imagers. We did not know how each detail in the scene was corrupted by the sampling process until the exact scene and its angular relationship to the sensor were specified. However, the *tendency* of the imager to produce visible display raster or corrupt scene details could be characterized.

For most practical, sampled imagers, the response function could be approximated by [23, 24]:

$$R_{sp}(\xi) \approx H_{post}(\xi)H_{pre}(\xi) + H_{post}(\xi)H_{pre}(\xi - \nu)e^{j\phi} + H_{post}(\xi)H_{pre}(\xi + \nu)e^{-j\phi} \quad (12.19)$$

where

H_{pre} = presample MTF; that is, the product of the optics MTF, detector MTF, and any MTF losses due to the atmosphere or line-of-sight jitter
H_{post} = postsample MTF; that is, the product of the electronics MTF, display MTF, and eye MTF
ξ = spatial frequency (cycle/mrad)
ν = sample frequency (cycle/mrad)
ϕ = sample phase.

Note that H_{TOT} in (12.13) is the product of H_{post} and H_{pre}.

The response function had two parts, the transfer term and the spurious response terms. The first term in (12.19) was the transfer response of the imager. This transfer response did not depend on sample spacing, and it was the only term that remained for small sample spacing. A sampled imager had the same transfer function as a nonsampled (that is, a very well sampled) imager.

However, a sampled imager always had the additional response terms, which we referred to as *spurious response*. These spurious response terms in (12.19) were filtered by the electronics, display, and eye MTFs in the same way that the transfer response was filtered. However, the position of the spurious response terms on the frequency axis depended on the sample spacing. If the sample spacing was large (the sample frequency was small), then the spurious response terms were close to the baseband in frequency space. In this case, the spurious response was difficult to filter out and might even overlap the baseband. If the sample spacing was small (the sample frequency is high), then the spurious response terms were far from the baseband in frequency space, and the spurious response was filtered out by the display and eye MTF.

The transfer and spurious response functions of a sampled imager could be calculated using (12.19). These response functions provided the Fourier transform of the baseband, desirable spatial image information, and the Fourier transform of the sampling artifacts, respectively. The sampled imager response function mathematically described the imaging behavior of the system. However, in predicting the effect of sampling on task performance, the response function had to somehow be condensed into a sampling-goodness metric for the sensor. Some generalizations had to be made and a goodness factor or factors calculated.

Two aggregate quantities were defined that proved useful in predicting how the spurious response of a sampled imaging system affected task performance. The utility of these quantities was discovered during experiments looking at the effect of sampling on target acquisition performance. The experiments are described in Refs. [25–27]. The two quantities were total integrated spurious response ratio, SR, as defined by (12.20), and out-of-band spurious response ratio, $SR_{out\text{-}of\text{-}band}$ as defined by (12.21) and (12.22).

$$SR = \frac{\int_{-\infty}^{\infty} \left|H_{post}(\xi)\right| [H_{pre}^2(\xi - \nu) + H_{pre}^2(\xi + \nu)]^{1/2}\, d\xi}{\int_{-\infty}^{\infty} \left|H_{post}(\xi) H_{pre}(\xi)\right| d\xi} \tag{12.20}$$

$$SR_{in\text{-}band} = \frac{\int_{-\nu/2}^{\nu/2} \left|H_{post}(\xi)\right| [H_{pre}^2(\xi - \nu) + H_{pre}^2(\xi + \nu)]^{1/2}\, d\xi}{\int_{-\infty}^{\infty} \left|H_{post}(\xi) H_{pre}(\xi)\right| d\xi} \tag{12.21}$$

$$SR_{out\text{-}of\text{-}band} = SR - SR_{in\text{-}band} \tag{12.22}$$

The results of perception experiments conducted by NVL showed that in-band aliasing (aliasing that occurred at frequencies of less than half the sample rate) did not degrade target identification performance, but out-of-band aliasing (such as visible display raster) degraded identification performance significantly. Aliasing had less impact on the recognition task than the identification task, but both in-band and out-of-band aliasing degraded recognition performance to some extent.

Based on these experiments and other results reported in the literature, it appeared that in-band aliasing had a strong affect on low-level discrimination tasks like hot spot detection; out-of-band aliasing has only a minor impact on these tasks. For high-level discrimination tasks such as target identification, however, out-of-band aliasing had a significant impact on performance, whereas in-band aliasing had a very minor effect. For intermediate-level discrimination tasks such as target or character recognition, both in-band and out-of-band aliasing had a moderate impact on performance.

The performance loss associated with sampling was modeled as an increased blur on the imagery. The blur increase was characterized as a function of total spurious response for the recognition task and as a function of out-of-band spurious response for the identification task. Using the Fourier similarity theorem, an increase in blur was equivalent to a contraction of the MTF, thus the term *MTF squeeze*. The squeeze for the recognition task was

$$SQ_{rec} = 1 - 0.32SR \tag{12.23}$$

where SR was defined by (12.20). The squeeze factor for the identification task was

$$SQ_{ID} = 1 - 2^* SR_{out\text{-}of\text{-}band} \tag{12.24}$$

where $SR_{out\text{-}of\text{-}band}$ was defined by (12.21) and (12.22). A squeeze for the detection task was not developed.

The spurious response was calculated independently in the horizontal and vertical directions, and the squeeze factor calculated. A new $H_{TOT}(f)$ was calculated:

$$H_{sqz}(f/SQ_{rec}) = H_{TOT}(f) \quad \text{for recognition} \tag{12.25}$$

or

$$H_{sqz}(f/SQ_{ID}) = H_{TOT}(f) \quad \text{for identification} \tag{12.26}$$

The MTF squeeze approach was an empirically derived method for imposing a penalty for undersampling. The penalty was not as severe as the half-sample rate limit imposed by FLIR92, but it gave performance below that of a well-sampled imager with identical presample and postsample transfer functions.

12.8 Other Improvements Prior to the Target Task Performance Metric

Johnson's target discrimination criteria were widely used for 25 years for two reasons: The resulting model was simple, and the model predictions were reasonable. However, this elegantly simple model did not quantitatively predict all aspects of target acquisition performance.

Johnson's original experiments were performed using high-contrast targets. Rosell performed experiments using noisy targets; he found that the Johnson/NVL model was progressively more optimistic as noise increased [21]. If Rosell's data were correct, then the NVL model predicted optimistically for poor weather, battlefield obscurants, noisy detectors, and other cases where noise or contrast limits performance. Rosell, Biberman, and others proposed alternatives to the NVL model to correct this behavior [25, 26]. Field experience showed that these alternative models did not predict range performance as well as the NVL model. Nonetheless, it was still widely believed that the NVL model lost accuracy under high-noise or low-contrast conditions. For example, the army "force-on-force" war game required (modulation) contrast to be more than 0.02 for target acquisition to occur. Without this limitation, the NVL model predicted that targets could be detected, recognized, or identified with a contrast as low as 0.002. Experiments at NVL showed that the Johnson criterion was a decent predictor of the influence of noise and blur on target identification, but that other image metrics predicted ID performance more accurately, especially when the images were noisy [27].

Another limitation of the previous models, discussed more fully in Refs. [26] and [28], is that the model was a robust predictor of sensor system range performance only for generic detection, recognition, and identification tasks. Targets were represented only by an average area and average target-to-background contrast. The model provided an average or statistical prediction for a group of observers attempting the task many times with a variety of target types represented (that is, a variety of target types being detected, recognized, or identified). The model was useful in comparing sensor systems, because it predicted target acquisition performance for an ensemble of target types and aspects. However, the model did not predict the probability of discriminating a specific vehicle. Even today, the current approach using the target task performance (TTP) metric applies to an ensemble of targets. However, target sets are defined for various sensor requirements.

A large number of improvements were made to the previous model including how signal processing was addressed. From 2000 to 2010, signal processing became much more important in the performance of imaging systems. For undersampled imagers, super-resolution was addressed. For wide dynamic range imagers, localized area processing (LAP) was addressed. A huge amount of effort was applied to signal processing and its effect on imager performance. Many of these techniques are still applicable to the TTP metric model provided later in this book.

12.9 The TRM3 Model

Wittenstein in Germany developed a model called TRM3 [28] using the Johnson approach but with a modification for the undersampled staring arrays. The impact of undersampling in the image of a four-bar target could readily be seen by simply observing the change in the image as a function of spatial frequency and phase. The spatial frequency was defined as line pairs per angular extent of the target, and the phase specifies the relative location of the target image to the detector array raster. These effects were obviously not independent of each other, but for each target, an optimal phase was found where the observer could see the maximum amplitude modulation in the image of the target. MRT measurements in the past utilized this variation with phase by allowing the observer to optimize the displayed image through target or system line-of-sight changes during the measurement process. For undersampled imagers, TRM3 addressed the problem of the MRT calculation's inability to predict beyond the half-sample rate of the sensor by replacing the MTF in the denominator of the MRT equation with an appropriately scaled term called the average modulation at optimum phase (AMOP). AMOP was the average signal difference in the image of the four-bar standard pattern, with the test pattern positioned at optimum phase. AMOP oscillated between the presample MTF and the bar modulation. Beyond a frequency of 0.8 times the sample rate or 1.6 times the half-sample frequency, the AMOP dropped rapidly to zero. The effect was to extend the prediction of the MRT beyond the half-sample rate and introduce a parameter called minimum temperature difference perceived (MTDP) to replace MRT in the range performance estimate and also in the lab to characterize an undersampled system.

The concept of MTDP followed from the observation of standard four-bar images as the pattern frequency passed the imager half-sample rate. The image changed from four bars to three, two, and finally one (was perceived by the observer). The phase of the target image on the detector array needed to be adjusted to observe this progression. The standard MRT measurement required that all four bars be resolvable by the observer during the measurement process. For use in the lab, the MTDP was defined as "the minimum temperature difference at which two, three, or four bars could be resolved in the image of the standard 4-bar test pattern by an observer, with the test pattern positioned at optimum phase. The optimum phase was the phase at which two, three, or four bars were best perceived." TRM3 used the standard definition of MRT as for adequately sampled imagers. The TRM3 approach model was implemented if the imager was considered undersampled, which

was defined as occurring when the prefilter MTF at the half-sample frequency was larger than 10%. The MTDP equation is given by

$$MTDP(f) = \frac{\frac{\pi}{2} SNR_{th} \Psi(f)}{AMOP(f)} \tag{12.27}$$

where SNR_{th} is the threshold signal-to-noise ratio and Ψ is the total system filtered noise. MTDP was used in the same manner as MRTD in range calculations.

For Germany, the FLIR92 model provided accurate performance predictions for first- and second-generation thermal imagers. However, the significant advances in the development of sensitive staring detectors led them to develop TRM3. A number of other countries adopted TRM3 as their sensor analysis and design model.

12.10 Triangle Orientation Discrimination

Another approach developed by Bijl in the Netherlands was the triangle orientation discrimination (TOD) threshold [29]. In TOD, the test pattern was a (nonperiodic) equilateral triangle in four possible orientations (apex up, down, left, or right), and the measurement procedure was a robust 4AFC (four alternative forced-choice) psychophysical procedure. In the procedure, the observer had to indicate which triangle orientation he saw, even if he was not sure (it was a forced-choice experiment). Variation of triangle contrast and/or size led to a variation in the percentage correct between 25% (complete guess) and 100%, and by interpolation the exact 75% correct threshold could be obtained. A complete TOD curve (comparable to an MRTD curve) was obtained by plotting the contrast thresholds as a function of the reciprocal of the triangle angular size. A detailed description of the measurement of a TOD curve is given in [29].

The TOD method had a large number of theoretical and practical advantages: it was suitable for undersampled and well-sampled EO and optical imaging systems in both the thermal and visual domains, it had a close relationship to real target acquisition, and the observer task was easy. The results were free from observer bias and allowed statistical significance tests. The method was implemented in standard MRTD test equipment with little effort, and the TOD curve could be used easily in a target acquisition (TA) model such as ACQUIRE. Two validation studies with real targets showed that the TOD curve predicted TA performance for undersampled and well-sampled imaging systems very well. The lab measurement and field performance appeared sound, but since the model was not separable in Cartesian coordinates, it could not be implemented in standard MTF and MRT approaches. A numerical model was developed due to the 2D nature of triangle frequency spectra.

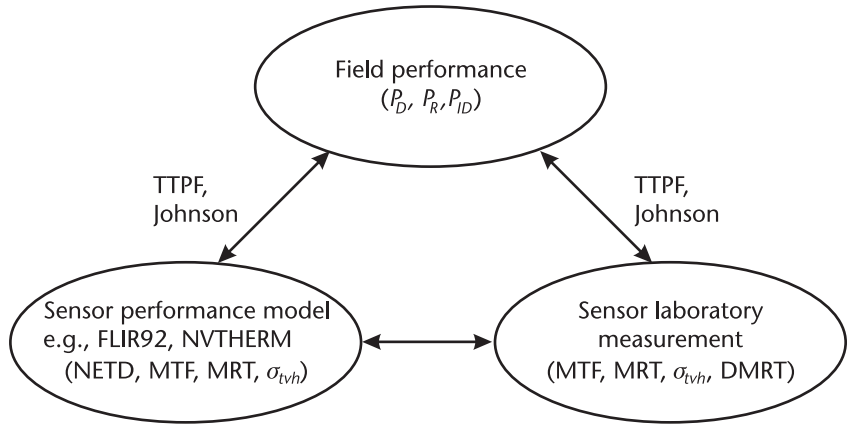


Figure 12.3 Modeling, testing, and performance triangle.

12.11 Imager Modeling, Measurement, and Field Performance

Sensor characterization is seen in three ways: theoretical models, field performance, and laboratory measurements (Figure 12.3). Theoretical models were developed that described sensor sensitivity, resolution, and human performance for the purpose of evaluating new conceptual sensors. Acquisition models and other models were developed to relate the theoretical models to field performance. This link allowed theoretical models to be converted to field performance quantities (e.g., probabilities of detection, recognition, and identification). Field performance was measured in the field so that the theoretical models could be refined and become more accurate with advanced sensor developments. Because field performance activities were so expensive, methods for the direct measurement of sensor performance were developed for the laboratory. Field performance testing of every infrared sensor built including buy-off, acceptance, and life-cycle testing was ridiculous and out of the question. Laboratory measurements of sensor performance were developed such that, given these measurements, the field performance of a system could be predicted. Sensor characterization programs required accurate sensor models, field performance estimates with acquisition models, and repeatable laboratory measurements.

12.12 Exercises

12.1 Given a 2.5K target contrast and a sensor cutoff frequency of 5 cycles per milliradian at that contrast, what is the range at which a tank can be identified? How about a jeep? Assume 100% atmospheric transmission. State all other assumptions.

12.2 What were the major differences in the original Ratches model and FLIR92?

12.3 What are the benefits of Triangle Orientation Descrimination compared to the classic periodic target measurements?

References

[1] Ratches, J., R. Vollmerhausen, and R. Driggers, "Target acquisition performance modeling of infrared imaging systems: Past, present, and future," *IEEE Sensors Journal, 2001,* Vol. 1, No. 1, 2001, pp. 31–40.

[2] Schade, O. "EO Characteristics of Television Systems," *RCA Review*, Vol. IX, Nos. 1–4, 1948.

[3] Sendall, R., and J. M. Lloyd, "Improved Specifications for Infrared Imaging Systems," *Proc. of IRIS*, Vol. 14, No. 2, pp. 109–129, 1970.

[4] Ratches, J. A., W. R. Lawson, L. P. Ober, R. J. Bergemann, T. W. Cassidy, et al., "NVL Static Performance Model for Thermal Viewing Systems," *USA Electronics Command Report Report # ECOM 7043*, 1973.

[5] Ratches, J. A., "Static Performance Model for Thermal Imaging Systems," *Optical Engineering*, Vol. 15, No. 6, 1976.

[6] Ratches, J. A., "Comparison of NVL Model and Four Contractor Models for Minimum Resolvable Temperature (MRT)," *USA Electronics Command Report Report # ECOM 7050*, 1976.

[7] Ratches, J. A., "NVL modeling; historical perspective," *Proc. of SPIE, Infrared Imaging Systems, Analysis, Modeling and Testing X*, Vol. 3701, 1999, pp. 1–12.

[8] Lawson, W. R., and James A. Ratches, "Modeling Detection or The Detection Game," *Proc. of IRIS Specialty Group on Targets, Backgrounds and Discrimination, Institute for Defense Analyses*, Arlington, VA, 1980.

[9] D'Agostino, J. A., et. al., "ACQUIRE Range Performance Model for Target Acquisition Systems Version 1 User's Guide," *USA CECOM Night Vision & Electronic Sensors Directorate Report*, Fort Belvoir, VA, 1995.

[10] Kneizys, F. X., E. P. Shettle, W. O. Gallery, J. H. Chetwynd, Jr., L. W. Abreu, et al., "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN 6," *AFGL-TR-83-0187*, Air Force Geophysics Laboratory, Hanscom AFB, MA, 1983.

[11] D'Agostino, J. A., et. al., "FLIR92 Thermal Imaging Systems Performance Model Analyst's Reference Guide," *Document UG5008993 USA CECOM Night Vision & Electronic Sensors Directorate*, Fort Belvoir, VA, 1993.

[12] Scott, L. B., and Lesley Condiff, "C2NVEO Advanced FLIR Systems Performance Model," *Proc. of SPIE Technical Symposium on Optical Engineering & Photonics in Aerospace Sensing*, Orlando, FL, 1990.

[13] D'Agostino, J. A., et. al., "FLIR92 Thermal Imaging Systems Performance Model User's Guide," *Document UG5008993 USA CECOM Night Vision & Electronic Sensors Directorate*, Fort Belvoir, VA, 1993.

[14] Driggers, R. G., R. Vollmerhausen, and K. Krapels, "Target Identification Performance as a Function of Temporal and Fixed Pattern Noise," *SPIE Proceedings*, Orlando, FL, 2000.

[15] Brown, J. C., C. Webb, P. Bell, R. Washington, and R. Riordan, "AUTOSPEC Image Evaluation Laboratory," *Proc. of SPIE, Infrared Imaging Systems*, Vol. 1488, 1991.

[16] Obert, L. P., J. A. D'Agostino, B. O'Kane, and C. Nguyen, "An Experimental Study of the Effect of Vertical Resolution on FLIR Performance," *Proc. of IRIS Specialty Group on Passive Sensors*, 1990.

[17] Howe, J. D., L. B. Scott, S. P. Pletz, J. D. Horger and J. S. Mark, "Thermal Model Improvement Through Perception Testing," *Proc. of IRIS Specialty Group on Passive Sensors*, 1989.

[18] O'Kane, B., M. Crenshaw, J. A. D'Agostino and D. Tomkinson, "Human Target Detection Using Thermal Devices," *Proc. of IRIS Specialty Group on Passive Sensors*, Vol. I, 1992, pp. 205–218.

[19] Vollmerhausen, R., "Incorporating Display Limitations into Night Vision Performance Models," *Proceedings of IRIS Specialty Group on Passive Sensors*, V2:11–31, 1995.

[20] Vollmerhausen, R., "Modeling the Performance of Imaging Sensors," in *EO Imaging: Systems and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[21] Rosell, F. A., and R. H. Wilson, "Recent Psychophysical Experiments and the Display Signal-to-Noise Ratio Concept," in *Perception of Displayed Information*, p. 167. L. M. Biberman, (ed), NY: Plenum, 1973.

[22] Kornfeld, G. H., and W. R. Lawson, "Visual Perception Models," *J. Opt. Soc. Am. Vol. 61*, 1971, pp. 811–820.

[23] Vollmerhausen, R., and R. Driggers, *Analysis of Sampled Imaging Systems, Tutorial Texts in Optical Engineering*, Vol. TT39, Bellingham, WA: SPIE Press, 2000.

[24] Vollmerhausen, R., "Display of Sampled Imagery," in *EO Imaging: Systems and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[25] Rosell, F. A., "Synthesis and Analysis of Imaging Sensors," in *EO Imaging: System Performance and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[26] Biberman, L., "Alternate Modeling Concepts," in *EO Imaging: System Performance and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[27] Vollmerhausen, R., R. Driggers, and M. Tomkinson, "Improved image quality metric for predicting tactical vehicle identification," *SPIE Proceedings*, Orlando, FL, 2000.

[28] Wittenstein, W., "Minimum temperature difference perceived – a new approach to assess undersampled thermal imagers," *Optical Engineering*, Vol. 38, No. 5, 1999, pp. 773–781.

[29] Bijl, P., and J. M. Valeton, "TOD, the alternative to MRTD and MRC," *Optical Engineering*, Vol. 37, No. 7, 1998, pp. 1984–1994.

CHAPTER 13

Contrast Threshold and TTP Metric

The ability to predict the probability of detecting, recognizing, or identifying a target as a function of range depends on how well the human eye sees; this is described by the contrast threshold function. Images are seen on displays that operate at a finite frame rate and because of sensor limitations there is frequently noise on the display. This motivates the requirement to define a contrast threshold function (CTF) of the eye with and without noise and a model that approximates this function. The work described here applies to any imaging systems regardless of whether the imaging system senses visible or infrared radiation, providing the imaging system produces a grayscale image that is viewed by the human eye.

We define an eye–display contrast function for any grayscale display (for example, a flat-panel or CRT display). Using this eye–display contrast function and the contrast of the target as seen on the display, a targeting task performance (TTP) metric is defined. It is useful to define a range-dependent visibility function V that depends on the TTP metric, target size L_{tgt}, and sensor–target range R. Using the V function and an empirically determined logistic function, the probability of performing the task as a function of range is then determined. The usual tasks modeled using the TTP metric are detection, recognition, and identification.

In Section 13.1 the contrast threshold function of the naked eye looking at a sinusoidal pattern on an ideal display, here defined as one without noise and infinite period (zero frame rate), is defined. A model that describes the performance of the eye with an ideal display is then described. In Section 13.2 the results of Section 13.1 are extended to the case of a display that has noise and a finite frame rate. Sections 13.1 and 13.2 describe attributes of the eye–display system that are purported to be a good approximation. In Section 13.3 we present experimental evidence and theory to support this contention. Here the equations of Section 13.2 are derived using dimensional analysis. (In Appendix 13A at the end of this chapter the equations of Section 13.2 are derived directly using properties of the eye.) In Section 13.4 *Mathematica* code is exhibited that implements the equations of Section 13.2 and comparison is made with the corresponding NV-IPM model calculations. Section 13.5 describes the TTP metric, logistic function, and V_{50} parameters and also outlines how the material in this chapter can be used to predict the performance of infrared, EO, and visible imaging systems.

13.1 Contrast Threshold Function of the Naked Eye

A sinusoidal display luminance pattern that is a function of an angle θ_h in the horizontal direction with luminance amplitude L_p and average luminance L_0 is generated on a computer display. As illustrated in Figures 13.1 and 13.2, the pattern is limited by angles w_h and w_v in the horizontal and vertical directions. Then

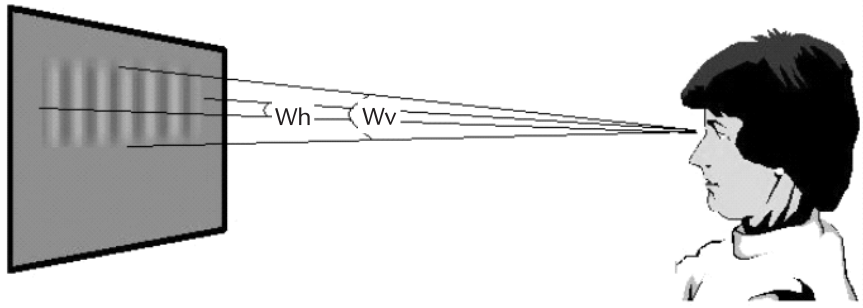


Figure 13.1 Two alternative forced choice experimental methods for determining contrast threshold function. The sinusoidal pattern is shown randomly at the top and bottom of the display and the observer is forced to decide if the pattern is on the top or bottom of the display.

$$L_{disp}(\theta_h) = L_0 + L_p \, sin(2\pi\xi\theta_h) \, rect\left(\frac{\theta_h}{w_h}\right) rect\left(\frac{\theta_v}{w_v}\right) \tag{13.1}$$

where ξ is measured in cycles per milliradian and, θ_h, θ_v, w_h and w_v are measured in milliradians.

The display contrast is defined by

$$C_{disp} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \tag{13.2a}$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum display luminance. From (13.1), $L_{\max} = L_0 + L_p$ and $L_{\min} = L_0 - L_p$. Substituting these results into (13.2a), we get

$$C_{disp} = \frac{Lp}{L_0} \tag{13.2b}$$

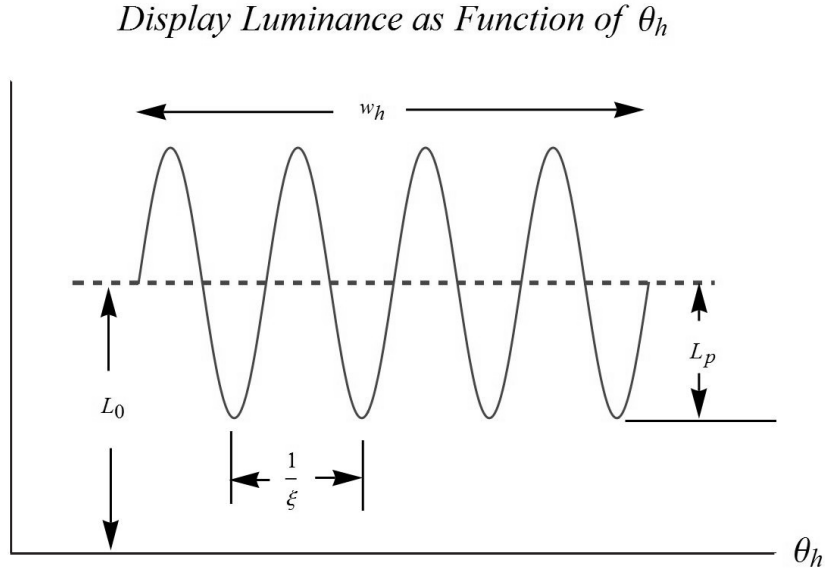


Figure 13.2 Illustrating L_0, L_p, ξ, and w_h.

That is, the contrast is the ratio of the luminance amplitude of the sinusoidal variation of the pattern shown in Figures 13.1 and 13.2 to the average luminance of the display.

From Figure 13.2 and (13.2b), we can see that C_{disp} is between 0 and 1. For a fixed value of L_p, since the luminance cannot be negative, the smallest value of L_0 is L_p and this implies that the maximum value of C_{disp} is 1. For a fixed L_0 the smallest value of L_p is zero, which implies that the smallest value of C_{disp} is 0.

The contrast threshold function and a technique for measuring it are described next. At a given horizontal spatial frequency ξ and display luminance level L_0, the contrast C_{disp} as seen on a computer display is made smaller and smaller by reducing L_p so that the modulated signal is barely visible. A zero contrast signal is shown randomly on either the lower or upper half of the screen in Figure 13.1 and a pattern with barely visible contrast is shown on the other half of the screen. The observer must decide if the modulated pattern is in the upper or lower fraction of the screen. The liminal contrast, at a given L_0 and spatial frequency ξ, is defined as the contrast where the observer correctly identifies which part of the screen had the modulated signal in three-fourths of the cases. After compensating for getting the correct answer by random guessing, this corresponds to the observer being able to detect the modulated signal 50% of the time. At a given luminance L_0, the spatial frequency ξ is allowed to vary and the experiment is repeated for several luminance values. Experimentally it is found that when w_h and w_v are greater than about 15 deg, the contrast threshold function is nearly independent of w_h or w_v but this is not the case when w_h or w_v are substantially less than 15 deg.

Barten [1] has generated smooth curves that summarize the results of the perception experiments illustrated in Figures 13.3 and 13.4. Barten's smooth curves were produced using the following equation:

$$CTF(\xi, L_0, w_h, n_{eye}) = min\left\{\frac{\sqrt{2}}{\sqrt{n_{eye}}}\left[a\,\xi e^{-b\xi}\sqrt{1+0.06e^{-b\xi}}\right]^{-1}, 1\right\} \tag{13.3}$$

where

$$a = \frac{540\left(1+\frac{0.7}{L_0}\right)^{-0.2}}{1+\frac{12}{w_h(1+0.333\xi)^2}}, \quad b = 0.3\left(\frac{100}{L_0}\right)^{0.15}$$

and $min\,(x_1, x_2)$ is the smaller of x_1, x_2. The *min* function guarantees that the *CTF* never exceeds 1.

The formula given by Barten [1] is for two eyes and does not have the factor $\sqrt{2/n_{eye}}$ or the *min* function. The factor $\sqrt{2/n_{eye}}$ that is given in (13.3) was included by Preece [9] and expresses the experimental result that *CTF* goes down when two eyes are used. This is consistent with the view that the signal is linearly proportional to the number of eyes used, and the noise is proportional to the square root of the number of eyes used so that the SNR is proportional to $\sqrt{n_{eye}}$. The *CTF* of the naked eye is inversely proportional to the SNR. The *min* function guarantees that *CTF* is never greater than 1.

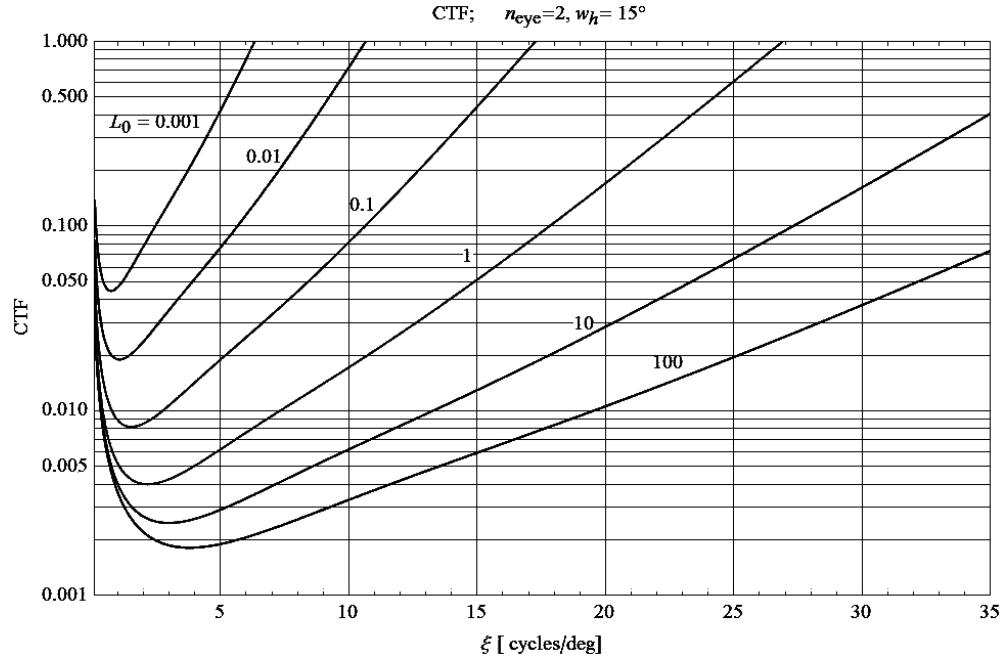


Figure 13.3 Contrast threshold function with two eyes, w_h = 15 deg and L_0 in units of candelas per square meter.

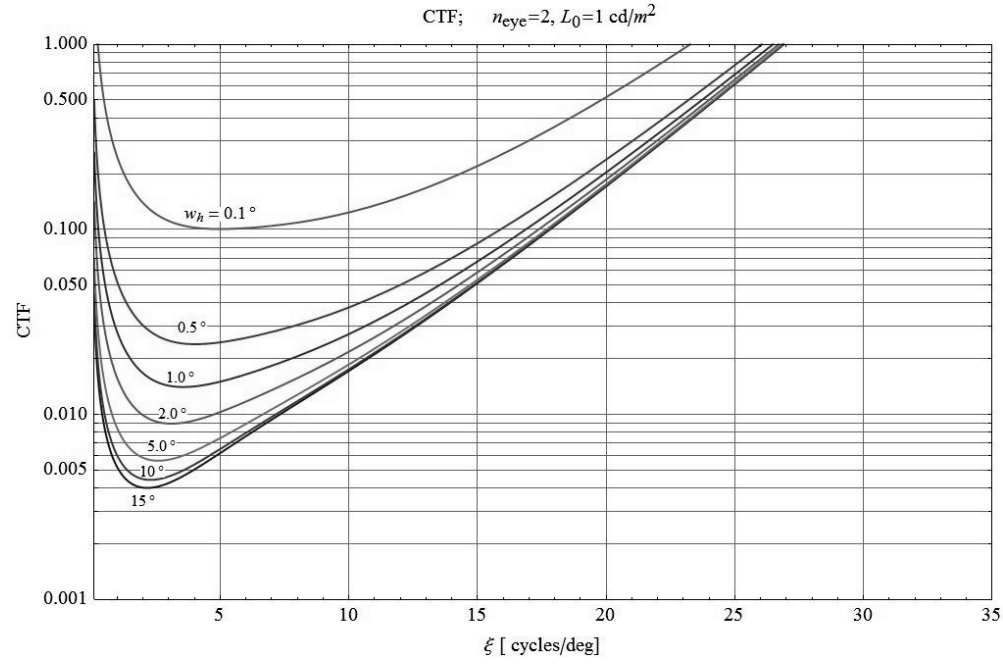


Figure 13.4 Contrast threshold function goes up as w_h goes down.

We use the subscript h and v to indicate the horizontal and vertical directions, respectively, and will subsequently use a t subscript to indicate a temporal variation.

Figures 13.3 and 13.4 illustrate how the contrast threshold function given by (13.3) varies as a function of spatial frequency with L_0 and w_h as parameters.

13.2 Contrast Threshold Function for the Eye–Display System

Section 13.1 described the performance of the human eye looking at a perfect display, that is, one with no noise and an infinite period (zero frame rate). In this section the results are extended to include displays with noise and a finite frame rate. Here the emphasis is on clearly stating what the results are and explaining why the results are reasonable. The results given in this section are derived in Section 13.3 using *dimensional analysis* [3, 4, 6, 13]. (Appendix 13A derives the results using properties of the eye–brain system.)

When the human visual system is coupled to an imaging device with noise and a finite frame rate, then the eye–display contrast threshold function has the following form [9]:

$$CTF_{eye\text{-}disp,h} = CTF\sqrt{1+\frac{\sigma^2_{eye\text{-}disp}}{\sigma^2_{eye}}} \tag{13.4}$$

where $\sigma^2_{eye\text{-}disp}$ is the perceived external noise variance of the brain–eye–display system, σ^2_{eye} is the effective internal brain–eye noise variance and CTF is the contrast threshold function in the absence of noise discussed in Section 13.1. This equation reasonably asserts that as the ratio of perceived display noise to the internal noise of the eye increases, then the eye–display CTF increases. Neither $\sigma^2_{eye\text{-}disp}$ nor σ^2_{eye} is directly available for measurement but, as will be seen, their values can be deduced from contrast threshold measurements performed in the presence of noise at varying frame rates. The term $CTF_{eye\text{–}disp,h}$ is expressed by [2]

$$CTF_{eye\text{-}disp,h} = CTF(\xi, L_0, w_h, n_{eye})\sqrt{1+\frac{\gamma^2 N^2_{disp,h,t} Q_{eye\text{-}disp,h}(\xi) Q_{eye\text{-}disp,t}}{L_0^2 \dfrac{1}{n_{eye}}\left(1+\dfrac{\beta^2 Q_{eye,t}}{L_0 A_{eff}}\right)}} \tag{13.5a}$$

Equations (13.4) and (13.5a) are the main results of this chapter. Comparing (13.4) and (13.5a), we identify

$$\sigma^2_{eye} = \frac{1}{n_{eye}}\left(1+\frac{\beta^2 Q_{eye,t}}{L_0 A_{eff}}\right) \tag{13.5b}$$

$$\sigma^2_{eye\text{-}disp} = \frac{\gamma^2 N^2_{disp,h,t} Q_{eye\text{-}disp,h}(\xi) Q_{eye\text{-}disp,t}}{L_0^2} \tag{13.5c}$$

As will soon become apparent, these can be thought of as dimensionless variances internal to the naked eye and the perceived variance of the eye viewing a display.

We now identify terms that appear in the square root of (13.5a). Here γ is a dimensionless calibration constant with an approximate value of 330; $N^2_{disp,h,t}$ is a noise power spectral density input to the display in the horizontal direction with units of $fL^2\ mr\ sec$; $Q_{eye\text{–}disp,h}(\xi)$ is an eye-display spatial bandwidth with units of mr^{-1} that depends on the spatial frequency ξ of the sinusoidal illuminance pattern shown on the display; $Q_{eye\text{–}disp,t}$ is a temporal bandwidth of the eye-display system with units of sec^{-1}; β is a calibration constant with an approximate value of $5.0\sqrt{Trolands \cdot s}$; $Q_{eye,t}$ is a temporal bandwidth associated with the naked eye with units of sec^{-1}; A_{eff} is the effective area of the eye's entrance pupil in mm^2; and L_0 is the average display luminance in fL.

Note the reasonableness of (13.5a). The retinal illuminance in Trolands is [11] the product of the average luminance L_0 in cd/m^2 and A_{eff} in units of mm^2. Since the dimensions of $L_0 A_{eff}$ are in Trolands, the second term in (13.5b) is dimensionless. Similarly both sides of (13.5c) are dimensionless. We conclude that (13.5a) is dimensionally consistent. Equation (13.5a) also has the right limiting behavior: when $N^2_{disp,h,t} / L_0^2 \approx 0$, then (13.5a) reduces to $CTF(\xi, L_0, w_h, n_{eye})$ as expected. Further evidence for the capabilities and limitations of (13.5a) is exhibited in the next section.

The reason A_{eff} is not the same as the geometric eye entrance pupil area is the Stiles-Crawford effect. Light passing near the periphery of the pupil is not as effective in exciting cone photoreceptors as are rays that enter near the center of the pupil [12].

To use (13.5a) to predict how well the eye will work with a particular display, we need to relate the quantities that appear in (13.5a) to measured input parameters and this is done in the remainder of this section.

To calculate A_{eff} in (13.5a), the pupil diameter is needed. The pupil diameter in millimeters, as given by Barten [2], depends on the average display luminance L_0 in fL and the field angle X_0 in degrees:

$$d_{pup} = 5 - 3\ tanh\left[0.4 log_{10}\left(L_0 3.426 \frac{X_0^2}{40^2}\right)\right] \tag{13.6}$$

We identify the field angle with the angle subtended by the display at the eye. Equation (13.6) is illustrated in Figure 13.5. As illustrated later in Figure 13.10, the pupil diameter also depends on the number of eyes used. The dark-adapted pupil diameter also depends on age [14].

The effective area of the pupil in square millimeters, which is useful for calculating the illuminance in the eye, is the geometric area of the entrance pupil multiplied by a correction factor that reflects the Stiles-Crawford effect [9]:

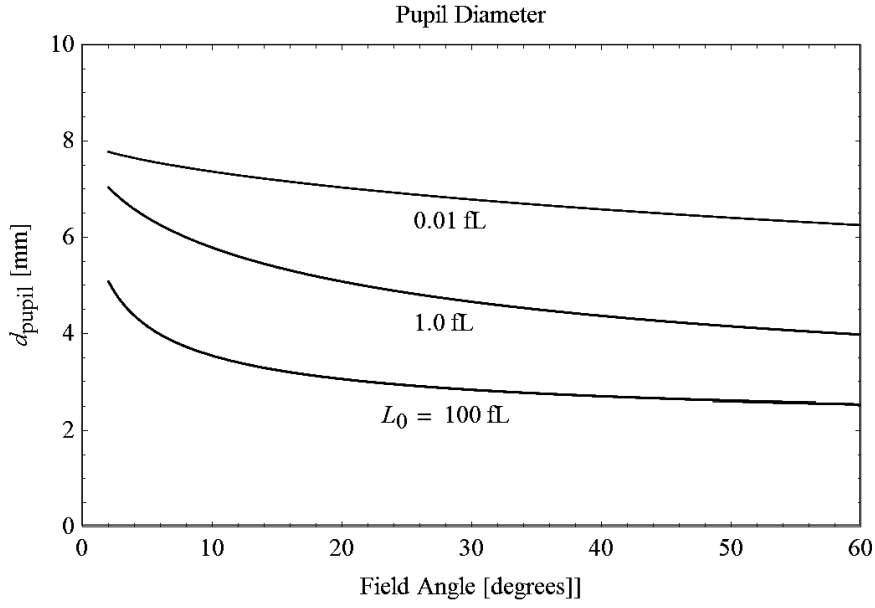


Figure 13.5 Pupil diameter as a function of field angle and luminance value.

$$A_{eff} = \frac{\pi}{4} d_{pup}^2 \left[1 - \left(\frac{d_{pup}}{9.7} \right)^2 + \left(\frac{d_{pup}}{12.4} \right)^4 \right] \tag{13.7}$$

Note that the correction term in square brackets is applicable to photopic and mesopic illuminance values and is appropriate for a typical soldier using tactical displays with a luminance of 10 fL. The value of the square bracket correction factor in (13.7) as a function of pupil diameter is illustrated in Figure 13.6.

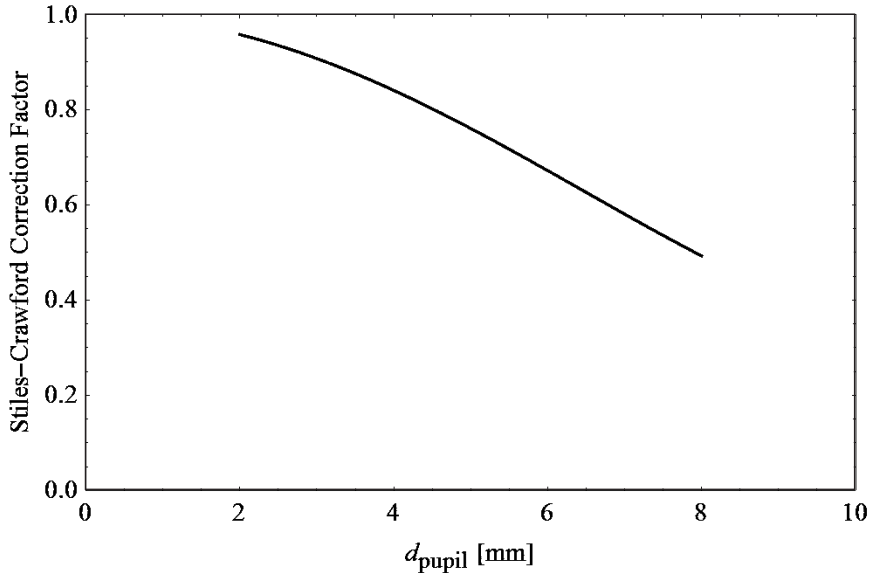


Figure 13.6 Stiles-Crawford correction factor as function of pupil diameter.

To calculate $Q_{eye,t}$, suppose the luminance of the source has a sinusoidal time dependence:

$$L_{source} = L_0(1 + sin2\pi\nu t) \tag{13.8}$$

For this source L_0 equals L_p, so the modulation of the source calculated using (13.2b) is one and at low spatial frequencies this is the modulation perceived by the eye. However, as temporal frequency ν increases the apparent modulation perceived by an observer decreases and has the form [9]

$$H_{eye,t}(\nu, \nu_{eye}) = \frac{\nu_{eye}}{\sqrt{\nu_{eye}^2 + (2\pi\nu)^2}} = \frac{1}{\sqrt{1 + 4\pi^2\left(\frac{\nu}{\nu_{eye}}\right)^2}} \tag{13.9a}$$

where ν_{eye}, a characteristic of the eye, is approximately independent of temporal frequency but does depend on luminance level. As illustrated in Figure 13.7, (13.9a) asserts that for í much less than ν_{eye} the eye perceives the full modulation, and for ν much larger than ν_{eye} the eye perceives no modulation.

It is convenient to define a time constant τ_{eye} for the eye in terms of ν_{eye}

$$\tau_{eye} \equiv \frac{2}{\nu_{eye}} \tag{13.9b}$$

The factor of 2 in (13.9b) is introduced to get a simpler expression for $Q_{eye,t}$. The noise equivalent bandwidth Q for a system with a transfer function $H(\nu)$ with a white noise source input is [5]

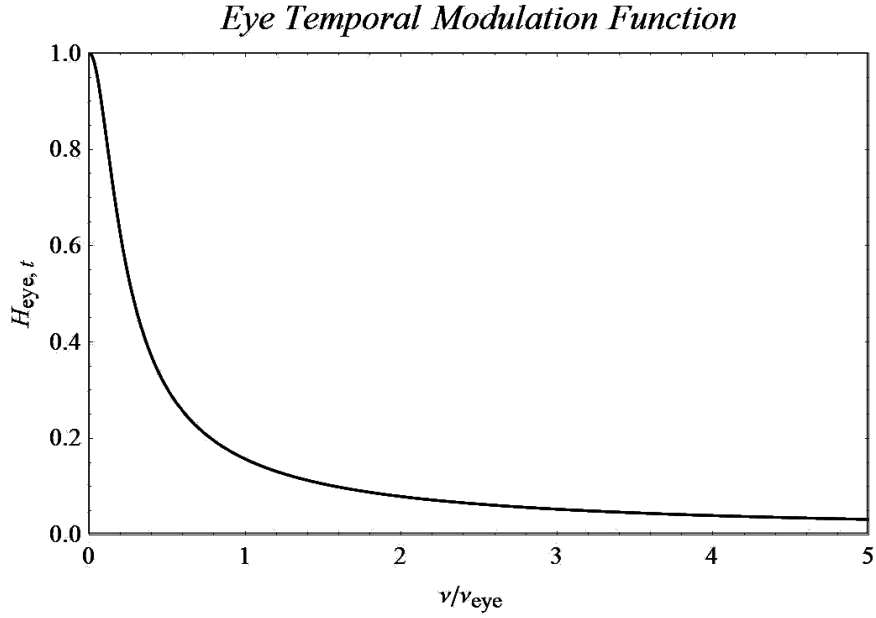


Figure 13.7 Eye temporal modulation function.

$$Q = \frac{1}{(|H_{\max}|)^2}\int_{-\infty}^{\infty}(|H(\nu)|)\,d\nu = \frac{2}{(|H_{\max}|)^2}\int_{0}^{\infty}(|H(\nu)|)\,d\nu \tag{13.10}$$

where $|H_{\max}|$ is the maximum of $|H(\nu)|$ and we exploited the evenness of $H(\nu)$ to convert the double-sided integral to a single-sided integral. Applying this result to $H_{eye,t}(\nu,\nu_{eye})$ given by (13.9a) we find that

$$Q_{eye,t} = 2\int_{0}^{\infty} H_{eye,t}(\nu,\nu_{eye})\,d\nu = \frac{\nu_{eye}}{2} = \frac{1}{\tau_{eye}} \tag{13.11}$$

To calculate $CTF_{eye-disp,h}$ using (13.5a), an explicit expression is needed for τ_{eye}. Contrast threshold data obtained in experiments done at NVESD support the view that the time constant of the eye increases with decreasing luminance and is described by the equation[9]

$$\tau_{eye} = 4\left[0.0192 + 0.0625\left(\frac{1.76}{L_0}\right)^{0.17}\right] \tag{13.12}$$

where τ_{eye} is the time constant of the eye in seconds and L_0 is the average luminance in fL. This equation is illustrated in the Figure 13.8.

Using (13.6), (13.7), (13.11), and (13.12), we can evaluate σ^2_{eye} using (13.5a). We seek an explicit expression for $\sigma^2_{eye\text{-}disp}$.

In (13.5c), $Q_{eye-\,disp,t}$ is a temporal bandwidth associated with the display and the eye. Using (13.10), we obtain

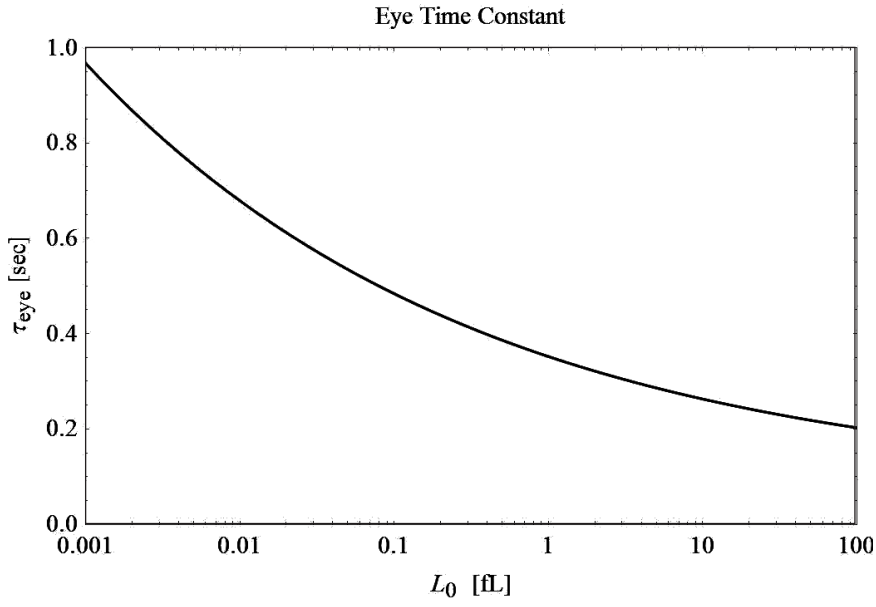


Figure 13.8 Eye time constant as a function of average luminance.

$$Q_{eye\text{-}disp,t} = 2\int_0^\infty (|\, H_{eye,t}(\nu, \tau_{eye}) H_{disp,t}(\nu, \tau_{disp})\,|)^2 \, d\nu \quad [\text{Hz}] \tag{13.13}$$

where $H_{eye,t}$ and $H_{disp,t}$ are the temporal transfer functions of the eye and display, and $Q_{eye\text{–}disp,t}$ is the noise equivalent temporal bandwidth of the eye–display system. An expression for $H_{eye,t}$ is given by (13.9a), (13.9b), and (13.12).

To evaluate $Q_{eye\text{–}disp,t}$, an explicit expression is needed for $H_{disp,t}$. The transfer function for the display depends on whether it is a flat-panel display or a CRT. For a flat-panel display, the impulse response is

$$h_{disp,t} = \frac{1}{\tau_{disp}} \, rect\left(\frac{t}{\tau_{disp}}\right) \quad -\infty < t < \infty \quad \text{for flat-panel display}$$

where τ_{disp} is the period the impulse is displayed. Taking the Fourier transform of the impulse response,

$$H_{disp,t}(\nu, \tau_{disp}) = sinc(\tau_{disp}\nu) \quad -\infty < \nu < \infty \quad \text{for flat-panel display} \tag{13.14}$$

For a CRT the modulation transfer function is given by an expression similar to that of (13.9a):

$$H_{disp,t}(\nu, \tau_{disp}) = \frac{\dfrac{2}{\tau_{disp}}}{\sqrt{\left(\dfrac{2}{\tau_{disp}}\right)^2 + (2\pi\nu)^2}} \quad -\infty < \nu < \infty \quad \text{for CRT display} \tag{13.15}$$

where τ_{disp} depends on the CRT phosphor. With $H_{\,disp,t}(\nu, \tau_{disp})$ given by (13.14) or (13.15), $Q_{\,eye\text{–}\,disp,t}$ can be calculated.

To evaluate (13.13), it is convenient to express (13.9a) in terms of τ_{eye}:

$$H_{eye,t}(\nu, \tau_{eye}) = \frac{\dfrac{2}{\tau_{eye}}}{\sqrt{\left(\dfrac{2}{\tau_{eye}}\right)^2 + (2\pi\nu)^2}} \quad -\infty < \nu < \infty \tag{13.16}$$

Using (13.14), (13.15), and (13.16) an explicit expression for $Q_{eye\text{–}disp,t}$ in (13.13) is found:

$$Q_{eye\text{-}disp,t}(\tau_{disp}, \tau_{eye}) = \begin{cases} \dfrac{1}{\tau_{disp}\left(1 + \dfrac{\tau_{eye}}{\tau_{disp}}\right)} & \text{for CRT display} \\ \dfrac{1}{\tau_{disp}}\left(1 + \left(e^{-2\frac{\tau_{disp}}{\tau_{eye}}} - 1\right)\dfrac{\tau_{eye}}{\tau_{disp}}\right) & \text{for flat-panel display} \end{cases} \tag{3.17}$$

Graphs of (13.17) are shown in Figure 13.9. Observe from (13.17) that as τ_{disp} approaches infinity $Q_{eye-disp,t}$ approaches zero for both the CRT and flat-panel displays.

The final term needed to evaluate (13.5a) is $Q_{eye-disp,h}(\xi)$, the equivalent spatial noise bandwidth of the eye–display system. Generalizing (13.10) to the spatial domain:

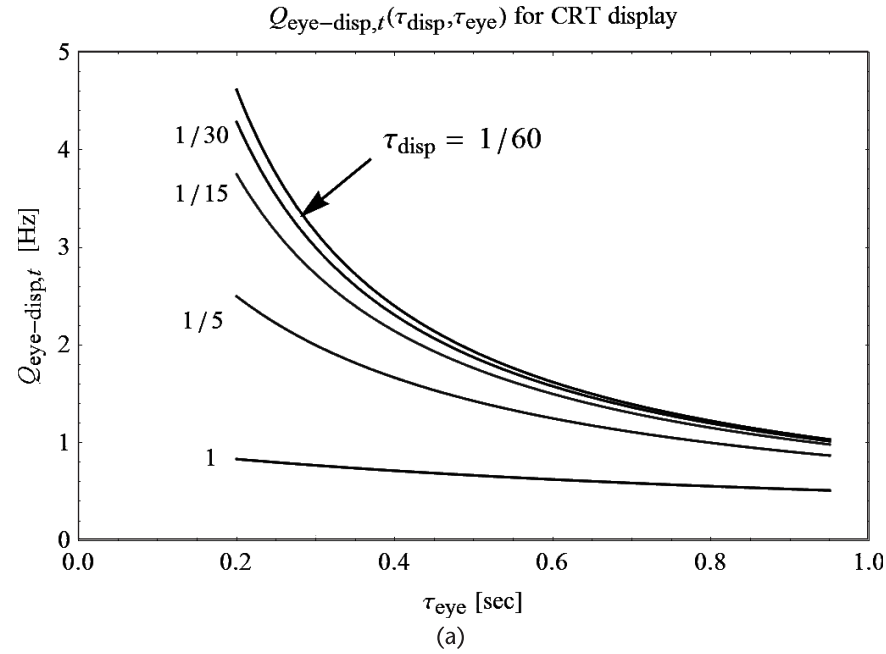


Figure 13.9 $Q_{eye-disp,t}$ for *CRT* and flat-panel displays.

$$Q_{eye\text{-}disp,h}(\xi) = \int_{-\infty}^{\infty} (|H_{disp}(\xi')H_{eye}(\xi')H_{chan}(\xi,\xi')|)^2 \, d\xi' \quad [\text{mrad}^{-1}] \tag{13.18}$$

Experimentally it is observed that $Q_{eye\text{–}disp,h}$ (ξ) is a function of the displayed spatial frequency. This necessitates the inclusion of the transfer function H_{chan} (ξ, ξ') for without this transfer function the noise equivalent bandwidth would be independent of spatial frequency.

To evaluate (13.18), explicit expressions are needed for H_{disp}, H_{eye}, and H_{chan} and these are now exhibited. The spatial transfer function of the display is different for flat-panel and CRT devices. For a CRT display the *psf* in the horizontal direction can be approximated by a Gaussian function:

$$psf(\theta_h) = \frac{1}{\sigma_{CRT}} e^{-\pi\left(\frac{\theta_h}{\sigma_{CRT}}\right)^2} \quad -\infty < \theta_h < \infty \tag{13.19a}$$

where θ_h is the angle in milliradians subtended at the eye in the horizontal direction, and σ_{CRT} is the value of θ_h where the *psf* is reduced by a factor of $e^{-\pi}$ from its maximum value. Here it has been assumed that the *psf* can be approximated by a Gaussian function. The display transfer function is the Fourier transform of the *psf*:

$$H_{disp}(\xi) = e^{-\pi(\sigma_{CRT}\,\xi)^2} \quad -\infty < \xi < \infty \quad \text{for CRT display} \tag{13.19b}$$

For a flat-panel display, the *psf* in the horizontal direction is well approximated by a *rect* function:

$$psf(\theta_h) = \frac{1}{\sigma_{FP}} rect\left(\frac{\theta_h}{\sigma_{FP}}\right) \quad -\infty < \theta_h < \infty \tag{13.20a}$$

with

$$\sigma_{FP} \equiv \frac{l_x}{d_{disp\text{-}obs}} \quad [\text{mrad}^{-1}]$$

where l_x is the horizontal dimension of a flat-panel display pixel in millimeters and $d_{disp\text{–}obs}$ is the distance between the observer and the display in meters. The Fourier transform of the *psf* yields the following flat-panel display transfer function:

$$H_{disp}(\xi) = sinc(\sigma_{FP}\xi) \quad -\infty < \xi < \infty \quad \text{for flat-panel display} \tag{13.20b}$$

The eye MTF, fit from data provided by Overington [8, 15], includes factors for refraction optics, retina, and tremor. The eye MTF depends on pupil diameter, which is a function of light level and the number of eyes used [15]:

$$d_{pup}(L_0, n_{eye}) = -9.011 + 13.23\, exp\left[-\frac{\log_{10} L_0}{21.082}\right] - (n_{eye} - 1)0.5 \tag{13.21a}$$

where the pupil diameter is given in millimeters and the light level is given in fL. Equation (13.21a) is illustrated in Figure 13.10.

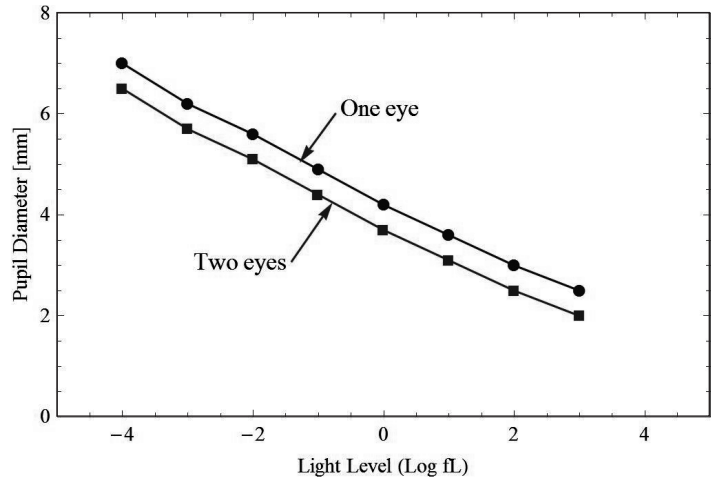


Figure 13.10 Pupil diameter as a function of light level for one and two eyes.

The eye MTF is given by [15]

$$MTF_{eye}(\xi, L_0, n_{eye}) = exp\left[-\left(\frac{f_e}{f_i}\right)^{e_0}\right] exp[-0.375\xi^{1.21}] exp[-0.4441\xi^2] \quad (13.21b)$$

with

$$e_0 \equiv \left(0.7155 + \frac{0.277}{\sqrt{d_{pup}}}\right)^2; \; f_i \equiv exp[3.663 - 0.0497 d_{pup}^2 \; log_{10} d_{pup}]; \; f_e \equiv 43.63\xi$$

where ξ and L_0 are given *in cycles/mr* and fL. Figure 13.11 illustrates the dependence of MTF_{eye} on luminance level and number of eyes used.
The channel *MTF* is given by[9,15]

$$H_{chan}(\xi, \xi') = exp\left[-2.2\left(log_e\left(\frac{|\xi'|}{\xi}\right)\right)^2\right] \quad 0 < \xi < \infty, \; -\infty < \xi' < \infty \quad (13.22)$$

In (13.22), think of ξ as the spatial frequency shown to the observer in Figure 13.1 and ξ' as the spatial frequency to which the brain is tuned. Equation (13.22), illustrated in Figure 13.12, is an empirical relationship that captures the observation that the spatial noise equivalent bandwidth depends on spatial frequency. Notice from Figure 13.12 and (13.22) that at the frequency where ξ' equals ξ, H_{chan} equals 1, indicating that the full *MTF* is obtained at the spatial frequency to which the brain is tuned.

Equations (13.19b), (13.20b), (13.21b), and (13.22) enable $Q_{eye\text{-}disp,h}(\xi)$ to be numerically calculated using (13.18).

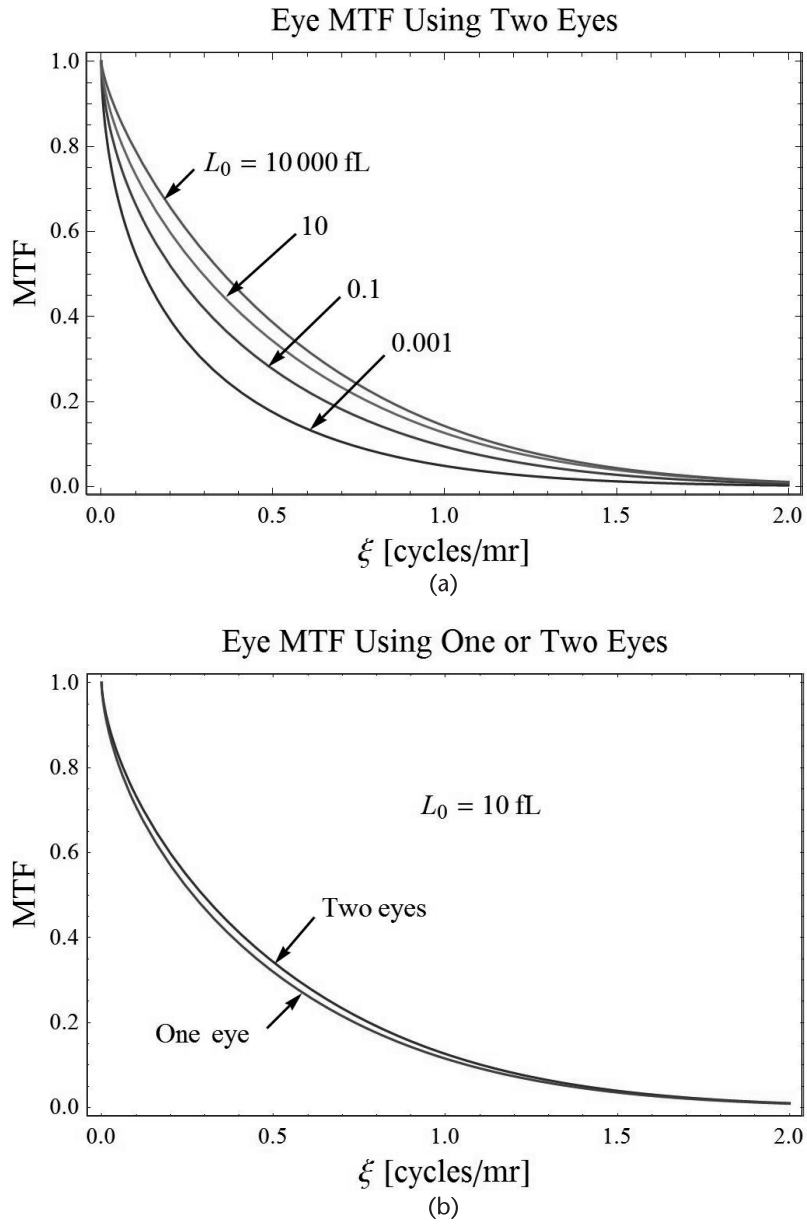


Figure 13.11 (a) Eye MTF for two eyes at different luminance levels. (b) Eye MTF for one or two eyes at 10 *fL*.

Using the equations in this section $CTF_{eye\text{-}disp,h}$ can be calculated for different noise power spectral densities, illuminance levels, and frame rates.

13.3 Validation of Eye–Display Contrast Threshold Model

In this section we present evidence that the equations in the previous two sections are valid over some range of input parameters and provide arguments to demonstrate that the equations in the previous section are reasonable. We suggest a modi-

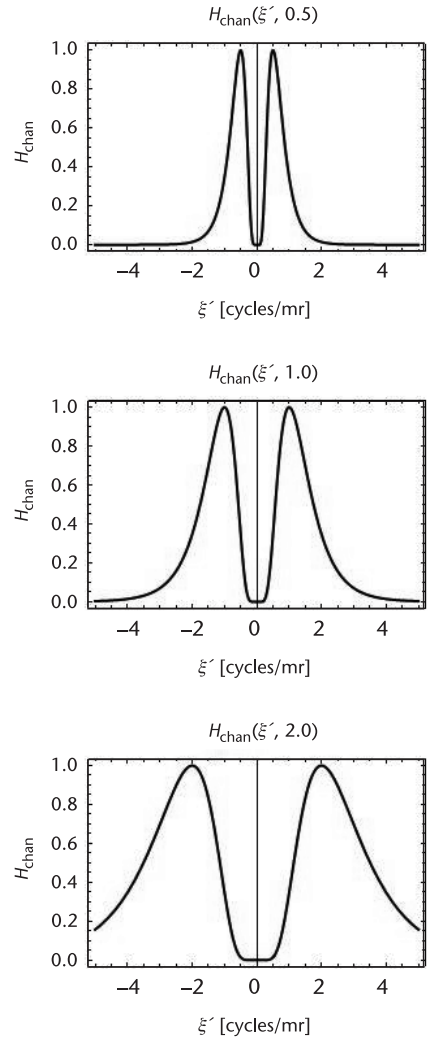


Figure 13.12 $h_{chan}(\xi,\xi')$ for ξ is 0.5, 1.0, and 2.0 cycles/mr.

fication of (13.5a) and an experimental protocol that may give better agreement between modeled and measured contrast threshold functions.

Preece [9] et al. have reported CTF experiments in which spatial frequency, average luminance level, input noise level, and frame rate were varied. The measured contrast threshold function CTF_{meas} was then compared with a modeled contrast threshold function CTF_{mod}. A comparison of CTF_{meas} versus CTF_{mod} is shown in Figure 13.13. In performing the experiments described in Figure 13.13 it was necessary to measure $N^2_{disp,h,t}$. This was done [9] in the following way. Known digital noise was generated and input into a computer. From the properties of the computer monitor, we knew how this noise mapped into fL on the display. This enabled the variance of the noise in fL^2 to be calculated. The angle in milliradians, subtended by a display pixel in the horizontal direction and the time interval in seconds the pixel noise was shown before the computer monitor was refreshed, is also known. Then $N^2_{disp,h,t}$ equals the calculated luminance variance multiplied by the product of the pixel subtended horizontal angle and display time.

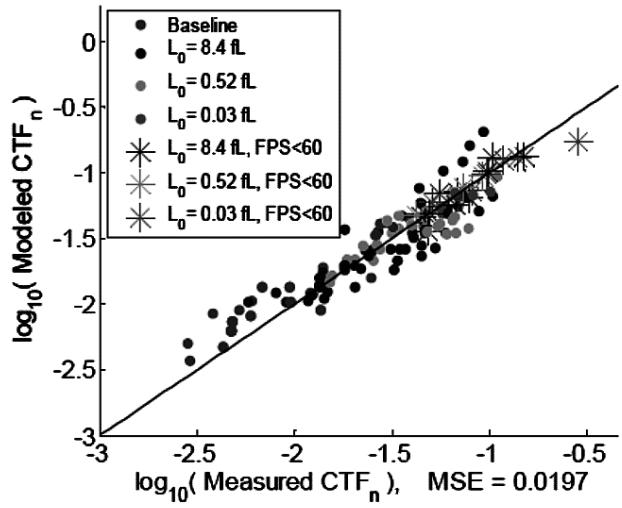


Figure 13.13 Comparison of measured and modeled contrast threshold function. Parameters that were varied include spatial frequency ξ, extent of sinusoidal display pattern w_h, input noise power spectral density $N^2_{disp,h,t}$, average luminance level L_0, and display time constant τ_{disp}.

In Figure 13.13 if the modeled and measured results were in complete agreement, all measurements would fall on the solid diagonal line. The measurements in Figure 13.13 are displayed in a *log–log* plot to accommodate the large range of the measured and modeled values. It is seen that the measured results are sometimes above and sometimes below the modeled results and that the measured and modeled contrast threshold results increase and decrease together. This supports the view that the equations of Section 13.2 are approximately correct.

The data collected for Figure 13.13 was collected by four observers (each with 20/20 vision) using the two-alternative forced choice method with 128 different threshold values. As indicated in the figure, the data include luminance values of 8.4, 0.525, and 0.0328 *fL*. The frame rates were set to static, 5, 10, 15, 30, and 60 frames per second. Experiments were done with one and two eyes viewing the display. The display viewing distance and size of the displayed modulation pattern w_h were changed so that w_h varied between 0.3 and 3.0 deg. The spatial frequencies shown on the monitor were allowed to take on the values 0.1, 0.2, 0.4, and 0.8 cycle/mr. The spatial noise patterns in the display were filtered using a lowpass filter and bandpass filter centered on the signal spatial frequency. Thus the modeled and measured results shown in Figure 13.13 cover a wide range of experimental conditions.

It is difficult to measure the contrast threshold function with high accuracy. Different observers with 20/20 vision differ in their CTF. Even for a particular observer, the ability to perceive a sinusoidal pattern at a given contrast varies depending on the fatigue level of the observer and also on the H_{chan} spatial frequencies to which the brain is tuned during the interval when the sinusoidal pattern is displayed. It is reasonable to expect that observers with 20/20 vision will have different liminal contrast in the presence of noise and a varying frame rate. Thus in Figure 13.3 each curve representing a given luminance level should be replaced by a band that represents the variance intrinsic to observers with 20/20 vision.

The theory of the contrast threshold function described in Section 13.2 is based on a published [9] eye–brain model. The reader is referred to this reference and Appendix 13A for a direct derivation of (13.5a). A critical review of the derivation allows the reader to assess the theoretical uncertainties in the results presented in Section 13.2.

Here we use dimensional analysis to investigate necessary relationships required among the variables of (13.5a). Our analysis assumes that the variables and a dimensional constant that characterizes the eye are sufficient to describe the eye–display contrast threshold function. As will soon be seen, *a virtue of dimensional analysis is that it allows one to deduce properties of a phenomenon without knowing the fundamental laws that govern the phenomenon*. Since the fundamental laws of how the eye–brain system works are not well understood, dimensional analysis is a suitable tool for finding necessary relationships between the variables of (13.5a).

To make the analysis done here accessible to readers with a modest background in dimensional analysis, we first illustrate the technique of dimensional analysis by modeling the period of a pendulum without invoking Newton's law of motion. Assume the period T of a pendulum depends on gravity g, the length of the pendulum L, and the mass M of the bob at the end of the pendulum. Clearly the period depends on g because if the acceleration of gravity were zero the pendulum would not oscillate. Intuitively, one expects the pendulum period to get longer with increasing length. The period might get shorter with increasing mass because of the increasing restoring force on the pendulum, but then again the period might get longer with increasing mass because a large mass is hard to accelerate. Without Newton's law of motion, the dependence on mass is hard to predict. Thus, it is reasonable to assume the period depends on g, L, and M and so we write

$$T = k\, g^{\epsilon_1} L^{\epsilon_2} M^{\epsilon_3} \tag{13.23a}$$

where k is a dimensionless constant and ϵ_1, ϵ_2, and ϵ_3 are constants that will make (13.23a) dimensionally consistent. Using the abbreviations *m, s,* and *kg* the dimensions of *T, g, L,* and *M* are, respectively, [*sec*], $[m \cdot s^{-2}]$, [*m*] and [*kg*]. The requirement that the dimensions on the left and right sides of (13.23) agree implies

$$\begin{aligned}[sec] &= [m \cdot s^{-2}]^{\epsilon_1} [m]^{\epsilon_2} [kg]^{\epsilon_3} \\ &= [kg]^{\epsilon_3} [m]^{\epsilon_1 + \epsilon_2} [sec]^{-2\epsilon_1}\end{aligned} \tag{13.23b}$$

The requirement that the dimensions on the left side of this equation be identical to those on the right hand side of the equation implies

$$\epsilon_3 = 0; \quad \epsilon_1 + \epsilon_2 = 0, \quad -2\epsilon_1 = 1 \tag{13.23c}$$

from which we deduce $\epsilon_1 = -\frac{1}{2}, \epsilon_2 = \frac{1}{2}$, and $\epsilon_3 = 0$. Using these results, (13.23a) becomes

$$T = k\sqrt{\frac{L}{g}} \tag{13.23d}$$

That is, using dimensional analysis we have found that the period of a pendulum does not depend on the mass M of the bob, but does depend on L and g as indicated in (13.23d). Dimensional analysis does not tell us the value of the dimensionless constant k. Instead we do one experiment in which we arbitrarily pick an L value and experimentally determine the period T. Then knowing T, L, and g, the constant k is found from (13.23d). The validity of (13.23d) can be conveniently determined by measuring the period for different values of M and L and less conveniently by varying g.

This example illustrates the savings in experimental work made possible with dimensional analysis. Without dimensional analysis, in the absence of a theory, if we guessed that the period depends on length, mass, and acceleration of gravity and we investigated the period for 10 values of length, 3 values of mass, and 3 values for the acceleration of gravity, we would need 90 measurements to establish a relationship between period, length, mass, and the acceleration of gravity. Using dimensional analysis, one measurement serves to determine the constant k, and a few other measurements confirm the correctness of the form given by (13.23d). Another benefit of dimensional analysis is that it provides confidence that the results will be valid outside the range of the independent variables that were allowed to vary in the experimental investigation.

The *Buckingham pi theorem* is fundamental to dimensional analysis. Any physical system can be described by a relationship [13]

$$\pi_1 = \Psi(\pi_2, \pi_3, \ldots, \pi_{N_p}) \tag{13.24a}$$

where the π_i are independent dimensionless combinations of physical variables and π_1 is the only dimensionless variable that contains the dependent physical variable. The number of independent physical variables is $N_p - 1$ (i.e., π_2, π_3,… π_{N_p}). These are dimensionless combinations that contain only *independent* physical variables. The Buckingham pi theorem takes as input the dimensions of physical variables that adequately describe a phenomenon and outputs the number N_p, the total number of *independent* dimensionless variables needed to characterize the phenomenon. As indicated in the pendulum example, the value in using dimensionless analysis is that the total number of dimensionless variables is usually less than the number of physical variables; hence, dimensional analysis has the potential to considerably simplify the experimental investigation of a physical relationship.

Since the Buckingham pi theorem is expressed in terms of *independent* dimensionless variables we need to define what the word *independent* means in this context. First we define a *dependent* dimensionless variable. The variable π_i is *dependent* on $\pi_1,\ldots,\pi_{i-1}$, π_{i+1}, …, π_{N_p} if constants $\in_i$ exist with the property

$$\pi_i = \pi_i^{\in_1} \ldots \pi_{i-1}^{\in_{i-1}} \pi_{i+1}^{\in_{i+1}} \ldots \pi_{N_p}^{\in_{N_p}} \tag{13.24b}$$

If there are no values of $\in_1,\ldots,\in_{i-1}$, $\in_{i+1}$, …, $\in_{N_P}$ that make (13.24b) true, then π_i is said to be *independent* of the remaining dimensionless variables. In the Buckingham pi theorem, all π_i are independent of one another.

Three examples will be used to illustrate the validity of (13.24a):

1. The electrical resistance R of a wire with cross-sectional area A, length L, and resistivity ρ is described by $R = (\rho\, L)/A$. In this example the physical phenomenon is described by four variables: R, ρ, L, and A, but is described by a single dimensionless variable $\pi_1 = (R\, A)/(\rho L)$.
2. The volume V of a rectangular box depends on the width w, height h, and length L of the box: $V = w\, h\, L$ but is described by a single dimensionless variable $\pi_1 = V/(w\, h\, L)$.
3. The distance h a ball thrown straight up as a function of time is given by $h(t) = -gt^2/2 + v_0 t + h_0$ where h_0 is the ball's initial height, v_0 is the ball's initial upward velocity, g is the acceleration of gravity, and t is time. This equation has five variables: h_0, v_0, g, t, and h but is described by three independent dimensionless variables: $\pi_1 = h/h_0$, $\pi_2 = g\, t^2/h_0$, and $\pi_3 = v_0\, t/h_0$.

We now show how dimensional analysis can be used to get an expression for the area A of a circle. The area of a circle can only depend on the radius r of the circle. Since A has the dimensions $[m^2]$ and the radius r has the dimension $[m]$ we can only make one dimensionless variable $\pi_1 = A/r^2$. Then (13.24a) implies that

$$\pi_1 = k \quad \Rightarrow \quad A = k\, r^2$$

where k is constant. When there is only one dimensionless variable Ψ in (13.24a) must be a constant since there are no remaining variables that would cause Ψ to change. Dimensional analysis does not tell us the value of the constant k. For this problem, the constant k is normally found by mathematical analysis but in other cases where an analytical approach is not practical or not possible, the constant can be found experimentally or by a combination involving mathematics, theory, and experiment.

The examples considered so far are trivial and were chosen to illustrate how dimensional analysis could be used in a familiar context. In the next example, a modification of an example taken from Szirtes [13], we use dimensional analysis to obtain a formula for computing the energy of an atomic explosion by observing the time evolution of the expanding hemispherical wavefront that ensues when a nuclear device is exploded near the earth's surface. Using physical insight (in dimensional analysis we always have to use physical insight or a preliminary analytical analysis to identify the dependent and independent variables), we expect the radius R of the wavefront to depend on time t, the amount of energy Q in the explosion, and the density ρ_0 of the air prior to the explosion. To find out how many independent dimensionless products are needed to describe this phenomenon, we look for coefficients ϵ_1, ϵ_2, ϵ_3, and ϵ_4 that will make a dimensionless variable π where

$$\begin{aligned} \pi &= R^{\epsilon_1} t^{\epsilon_2} \rho_0^{\epsilon_3} Q^{\epsilon_4} \\ &= [m]^{\epsilon_1} [sec]^{\epsilon_2} [kg \cdot m^{-3}]^{\epsilon_3} [kg \cdot m^2 \cdot s^{-2}]^{\epsilon_4} \\ &= [kg]^{\epsilon_3 + \epsilon_4} [m]^{\epsilon_1 - 3\epsilon_3 + 2\epsilon_4} [sec]^{\epsilon_2 - 2\epsilon_4} \end{aligned} \tag{13.25a}$$

from which we conclude

$$\begin{aligned} \epsilon_3 + \epsilon_4 &= 0 \\ \epsilon_1 - 3\epsilon_3 + 2\epsilon_4 &= 0 \\ \epsilon_2 - 2\epsilon_4 &= 0 \end{aligned} \tag{13.25b}$$

Equation (13.25a) represents three equations in four variables. This set of equations is readily solved for ϵ_2,ϵ_3, and ϵ_4 in terms of ϵ_1:

$$\epsilon_2 = -\frac{2}{5}\epsilon_1; \quad \epsilon_3 = \frac{1}{5}\epsilon_1; \quad \epsilon_4 = -\frac{1}{5}\epsilon_1 \tag{13.25c}$$

Thus π in (13.25a) is given by

$$\pi = \left(R^1 t^{-\frac{2}{5}} \rho_0^{\frac{1}{5}} Q^{-\frac{1}{5}} \right)^{\epsilon_1} \tag{13.25d}$$

Since the power of a dimensionless product is not independent, the value of ϵ_1 is arbitrary and in the interest of simplicity, we set ϵ_1 equal to 1. This problem is characterized by a single dimensionless product and from (13.24a) we conclude that

$$R^5 \sqrt{\frac{\rho_0}{Q t^2}} = k \tag{13.25e}$$

Experimentally it is found that nuclear explosions are described by (13.25e) and that k is approximately equal to 1 independent of the values of R, ρ_0, Q, and t, confirming that the single variable π describes the phenomenon. Solving (13.25e) for Q, we get

$$Q = \rho_0 \frac{R^5}{t^2} \tag{13.25f}$$

By measuring the radius of the wave-front at a particular time and knowing the air density prior to the explosion, the energy of the explosion can be calculated. In this example dimensional analysis reduced a complicated modeling problem that involved four variables to a single dimensionless variable that could be investigated with relative ease with a small number of measurements. The success of the program relied critically on identifying the variables that adequately describe the phenomena.

In (13.25c), the equations were solved in terms of ϵ_1 but the physical result is independent of the chosen variable. To illustrate this, solve (13.25c) in terms of ϵ_2. Doing this we find

$$\epsilon_1 = -\frac{5}{2}\epsilon_2; \quad \epsilon_3 = -\frac{1}{2}\epsilon_2; \quad \epsilon_4 = \frac{1}{2}\epsilon_2 \tag{13.25g}$$

which results in the dimensionless variable

$$\pi = \left(R^{-\frac{5}{2}} t \rho_0^{-\frac{1}{2}} Q^{\frac{1}{2}} \right)^{\epsilon_2} \tag{13.25h}$$

The value of ϵ_2 is arbitrary and if we choose it to be –2/5 we get the same dimensionless variable we obtained in (13.25d). The reader may check that (13.25f) is obtained regardless of which set of three variables are solved for in (13.25b). Dimensional analysis is discussed in depth in books by Szirtes [13], Bridgeman [4], Langhaar [6], and Bernstein [3].

With this background we now use dimensional analysis to derive (13.4). Using physical insight we expect that $CTF_{eye-disp,h}$ looking at a noisy display to be related to the contrast threshold function of the eye looking at a perfect display *CTF*, a dimensionless variance σ^2_{eye} that represents an internal noise floor of the eye–brain system, and a dimensionless variable $\sigma^2_{eye-disp}$ that represents the perceived noise variance of the eye–brain system as it views an image as shown in Figure 13.1. One approach is to say there are four dimensionless variables: $\pi_1 = CTF_{eye=disp,h}$, $\pi_2 = CTF$, $\pi_3 = \sigma^2_{eye}$, and $\pi_4 = \sigma^2_{eye-disp}$. Then we would assume that (13.24a) is valid with four variables. A more inspired approach is to realize that $CTF_{eye-disp,h}$ will be expressed in terms of *CTF*, and the ratio $CTF_{eye-disp,h}/CTF$ will depend on the ratio $\sigma^2_{eye-disp}/\sigma^2_{eye}$. Thus, we define two dimensionless variables:

$$\pi_1 \equiv \frac{CTF_{eye-disp,h}}{CTF}; \quad \pi_2 \equiv \frac{\sigma^2_{eye-disp}}{\sigma^2_{eye}} \tag{13.26}$$

and (13.24a) implies there exists a Ψ such that

$$\pi_1 = \Psi(\pi_2) \quad \Rightarrow \quad CTF_{eye-disp,h} = CTF\, f_1(\sigma^2_{eye-disp}/\sigma^2_{eye}); \quad f_1(0) = 1 \tag{13.27a}$$

where f_1 is a function we have yet to determine. We expect $CTF_{eye\text{-}disp,h}$ will equal *CTF* when $\sigma^2_{eye-disp}/\sigma^2_{eye} \ll 1$, which implies that f_1 (0) = 1. Assuming f_1 can be represented as a power function, (13.27a) becomes

$$CTF_{eye-disp,h} = CTF\left(1 + \frac{\sigma^2_{eye-disp}}{\sigma^2_{eye}}\right)^{\alpha_1} \tag{13.27b}$$

where α_1 is a constant.

If $\sigma^2_{eye\text{-}disp}$ and σ^2_{eye} were directly accessible parameters, then at last three approaches for determining $CTF_{eye-disp,h}$ are possible: 1) If we consider (13.27a) to be a fundamental, then with sufficiently good experimental data f_1 can be from the data. 2) If we consider (13.27b) to be a sufficiently good approximation, then we can determine a_1 from the data. 3) The third approach is to guess or use the analysis given in Appendix 13A to realize that α_1 = 1/2. With α_1 = 1/2, (13.27b) is (13.4).

We now use dimensional analysis to derive (13.5b). We expect σ^2_{eye} to increase as the temporal bandwidth $Q_{eye,t}$ of the eye increases. Further progress is facilitated by a brief review of radiometry and photometry. The unit of light power is the *lumen*; it is the analog of the *watt* in radiometry. The unit of irradiance for visible radiation is *illuminance*. Irradiance has the units of watts per square meter, whereas illuminance has the units of lumens per square meter. There is a separate unit for illuminance in the eye called the *Troland*. A *Troland* is [11] the luminance in *candelas per square meter* multiplied by the effective pupil area in square millimeters. We expect that σ^2_{eye}

will depend on the illuminance in the eye measured in Trolands. With L_0 the incident luminance on the eye in *candelas per square meter* and A_{eff} the effective pupil area in square millimeters, we expect that σ^2_{eye} will depend on the product $L_0\, A_{eff}$, which has units of Trolands. To express σ^2_{eye} in terms of a dimensionless variable π, we assume σ^2_{eye} depends on a dimensional constant β_1 with units of *Troland* $\cdot$ *s*. We now identify the dimensionless variables that characterize the problem:

$$\pi_1 = \sigma^2_{eye}; \qquad \pi_2 = \frac{\beta_1 Q_{eye,t}}{L_0 A_{eff}} \tag{13.28}$$

It is easy to confirm that π_2 is a dimensionless variable: $Q_{eye,t}$ has the dimensions of sec^{-1}, $L_0\, A_{eff}$ has the units of *Troland*, and since β_1 has units *Troland* $\cdot$ *s* we conclude that π_2 is dimensionless. Thus a function Ψ exists with the property

$$\pi_1 = \Psi(\pi_2) \quad \Rightarrow \quad \sigma^2_{eye} = f_2\left(\frac{\beta_1 Q_{eye,t}}{L_0\ A_{eff}}\right); \quad f_2(0) = \alpha_2 \neq 0 \tag{13.29}$$

where f_2 is a function we need to determine. If L_0 is very large, we would expect σ^2_{eye} to approach an asymptotic value $\alpha_2 \neq 0$. So even though we do not know f_2, physical considerations enable us to conclude that $f_2(0) = \alpha_2$.

Why do we expect the asymptotic value of σ^2_{eye} to approach a constant as L_0 increases? In Figure 13.3, as L_0 gets larger, the minimum in the contrast threshold function approaches a constant finite value and this minimum value is associated with the finite minimum value of σ^2_{eye} with increasing values of L_0. A reasonable guess is

$$\sigma^2_{eye} = \alpha_2\left(1 + \frac{\beta_1\, Q_{eye,t}}{L_0\, A_{eff}}\right)^{\alpha_3} \tag{13.30a}$$

where α_2 and α_3 are to be determined by experiment.

In the analysis of σ^2_{eye} that has been done so far, we have neglected the number of eyes used because often two eyes are used and we did not expect the number of eyes to be a major factor influencing σ^2_{eye}. If we allow the number of eyes to vary, we expect σ^2_{eye} to be smaller with two eyes than with one eye. Modifying (13.30a) to account for one or two eyes, we get

$$\sigma^2_{eye} = \frac{\alpha_2}{n_{eye}}\left(1 + \frac{\beta_1\, Q_{eye,t}}{L_0\, A_{eff}}\right)^{\alpha_3} \tag{13.30b}$$

Perception experiments [9] support the view $\alpha_2 = \alpha_3 = 1$. Then (13.30b) is identical to (13.5b) if we identify β_1 with β^2.

We now use dimensional analysis to derive (13.5c). We expect $\sigma^2_{eye-disp}$ to get larger with a larger temporal bandwidth $Q_{eye-disp,t}$ and with a larger spatial bandwidth $Q_{eye-disp,h}$. If the input noise power density spectrum $N^2_{disp,h,t} = 0$, then $\sigma^2_{eye-disp}$ is expected to be zero. The units of $Q_{eye-disp,t}$ are s^{-1}; the units of $Q_{eye-disp,h}$ are mr^{-1}; the units of $N^2_{disp,h,t}$ are $fL^2 \cdot mr \cdot s$. To create a dimensionless unit we need to introduce the ambient luminance L_0 in units of *fL*. The problem is then characterized by two dimensionless variables:

$$\pi_1 = \sigma^2_{eye-disp}; \quad \pi_2 \equiv \frac{N^2_{disp,h,t}\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2} \tag{13.31}$$

Thus a function Ψ exists such that

$$\pi_1 = \Psi(\pi_2) \quad \Rightarrow \quad \sigma^2_{eye-disp} = f_3\left(\frac{N^2_{disp,h,t}\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2}\right); \quad f_3(0) = 0 \tag{13.32a}$$

where f_3 is be determined. The condition that $f_3\,(0) = 0$ follows from the realization that when $N^2_{disp,h,t} = 0$, then $\sigma^2_{eye-disp} = 0$. A reasonable approximation is

$$\sigma^2_{eye-disp} = \alpha_5\left(\frac{N^2_{disp,h,t}\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2}\right)^{\alpha_4} \tag{13.32b}$$

If $\alpha_4 = 1$ and $\alpha_5 = y^2$, then (13.32b) is identical to (13.5c).

We now use dimensional analysis to obtain (13.5a). From (13.27a)

$$CTF_{eye-disp,h} = CTF\, f_1\left(\sigma^2_{eye-disp}/\sigma^2_{eye}\right); \quad f_1(0) = 1 \tag{13.33a}$$

From (13.29) and (13.30b)

$$\sigma^2_{eye} = \frac{1}{\eta_{eye}} f_2\left(\frac{\beta_1\ Q_{eye,t}}{L_0\ A_{eff}}\right); \quad f_2(0) = \alpha_2 \neq 0 \tag{13.33b}$$

From (13.32a)

$$\sigma^2_{eye-disp} = f_3\left(\frac{N^2_{disp,h,t},\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2}\right); \quad f_3(0) = 0 \tag{13.33c}$$

If it is true that the ratio $CTF_{eye-disp,h}\,/\,CTF$ depends only on the ratio $\sigma^2_{eye-disp}/\sigma^2_{eye}$, then (13.33a) is the most general relationship that is allowed. If σ^2_{eye} depends on $Q_{eye,t}$, L_0, A_{eff} and n_{eye} then there is good reason to think that (13.33b) is the most general allowed relationship. Similarly if $\sigma^2_{eye-disp}$ depends on $N^2_{disp,h,t}$, $Q_{eye-disp,h}$, $Q_{eye-disp,t}$, and L_0 then (13.33c) is the most general form for this relationship.

If it is assumed that f_1, f_2, and f_3 are expressible as powers of the dimensionless variables, then combining (13.27b), (13.32b), and (13.30b), we get

$$\begin{aligned} CTF_{eye-disp,h} &= CTF\left(1 + \frac{\alpha_5\left(\dfrac{N^2_{disp,h,t}\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2}\right)^{\alpha_4}}{\dfrac{\alpha_2}{\eta_{eye}}\left(1 + \dfrac{\beta_1\ Q_{eye,t}}{L_0\ A_{eff}}\right)^{\alpha_3}}\right)^{\alpha_1} \\ &= CTF\left(1 + \frac{\gamma^2\left(\dfrac{N^2_{disp,h,t}\ Q_{eye-disp,h}\ Q_{eye-disp,t}}{L_0^2}\right)}{\dfrac{1}{\eta_{eye}}\left(1 + \dfrac{\beta^2\ Q_{eye,t}}{L_0\ A_{eff}}\right)}\right)^{1/2} \end{aligned} \tag{13.34}$$

where we have set $\alpha_1 = 1/2$, $\alpha_2 = \alpha_3 = \alpha_4 = 1$, and $\alpha_5 = \gamma^2$, and identify β_1 as the constant β^2. Equation (13.34) is in agreement with (13.5a).

Figure 13.13 earlier in the chapter shows that although measured and modeled eye–display contrasts increase or decrease together, it also indicates there is room for improvement. If the contrast threshold function does indeed depend on the variables indicated in Section 13.2, then the dimensional analysis provides a method for getting better agreement between modeled and measured contrast threshold functions. We accept the validity of (13.35a):

$$CTF_{eye-disp,h} = CTF(\xi, L_0, w_h, \eta_{eye})\sqrt{1 + \frac{\sigma^e_{eye-disp}}{\sigma^2_{eye}}} \tag{13.35a}$$

Equation (13.35b) then follows from (13.33b) and (13.33c):

$$CTF_{eye-disp,h} = CTF(\xi, L_0, w_h, \eta_{eye})\sqrt{1 + \frac{f_3\left(\dfrac{N^2_{disp,h,t}\, Q_{eye-disp,h}\, Q_{eye-disp,t}}{L_0^2}\right)}{\dfrac{1}{\eta_{eye}} f_2\left(\dfrac{\beta_1\, Q_{eye,t}}{L_0\, A_{eff}}\right)}} \tag{13.35b}$$

The function f_3 can be found in the following way: $CTF(\xi, L_0, w_h, n_{eye})$ is known from Barten's (13.3) and $CTF_{eye-disp,h}$ can be measured experimentally. By varying $N^2_{disp,h,t}$ the argument of f_3 is varied while the argument of f_2 remains constant. This enables f_3 to be measured as a function of its single argument. The argument of f_2 can be changed by varying L_0 and if $N^2_{disp,h,t}$ is allowed to vary so $N^2_{disp,h,t}/L_0^2$ remains constant, then the function f_2 can be determined by measuring $CTF_{eye-disp}$. This approach is expected to give a better agreement between modeled and measured contrast threshold functions and is also expected to be useful for more accurately modeling the performance of imaging sensors.

Dimensional analysis provides a method for exploring other models whose validity can be experimentally investigated. For example, suppose the dimensions of β^2 which characterizes the performance of the eye were changed from *Troland·s* to *Troland*^{2}*·s*. Then (13.35b) would change to

$$CTF_{eye-disp,h} = CTF(\xi, L_0, w_h, \eta_{eye})\sqrt{1 + \frac{f_3\left(\dfrac{N^2_{disp,h,t}\, Q_{eye-disp,h}\, Q_{eye-disp,t}}{L_0^2}\right)}{\dfrac{1}{\eta_{eye}} f_2\left(\dfrac{\beta_1\, Q_{eye,t}}{(L_0\, A_{eff})^2}\right)}} \tag{13.35c}$$

The program for determining the functions f_3 and f_2 outlined for (13.35b) applies equally well to (13.35c). It would be interesting to see which model better describes the experimental data.

It is instructive to see if (13.5a) gives the correct behavior when $\tau_{disp} \gg \tau_{eye}$. From (13.17), it is apparent that under these conditions $Q_{eye-disp,t} \approx 1/\tau_{disp}$ for both *CRT* and flat-panel displays. From the discussion in the beginning of Section 13.3 describing the measurement of $N^2_{disp,h,t}$, it is apparent that $N^2_{disp,h,t}$ is proportional to τ_{disp}. Thus, when $\tau_{disp} \gg \tau_{eye}$, the product $N^2_{disp,h,t}\ Q_{eye-disp,t}$ is constant independent of τ_{eye} and τ_{disp}. It is expected that $CTF_{eye-disp,h}$ would not depend on τ_{eye} or τ_{disp} for an observer looking at a static image and this is just what is predicted by (13.5a).

We know that the *CTF* measured using the technique of Figure 13.1 depends on w_v as well as w_h. Equation (13.3) needs to be generalized so it includes w_v. Approaches for doing that are proposed in Exercise 13.10. The validity of that formula needs to be confirmed by experiment.

Experiments need to be done to better characterize τ_{eye} and d_{pup}. The reader should note that the value of τ_{eye} given in (13.12) and illustrated in Figure 13.8 is a factor of 4 larger than the value of τ_{eye} used in the previous NVESD thermal imaging model. As shown in Figure 13.14, for a typical display with a field angle of 20 deg, (13.6) gives a larger pupil diameter than (13.21a) using one or two eyes. This supports the view that experiments need to be done to better characterize τ_{eye} and d_{pup}.

13.4 Eye–Display Contrast Threshold Model

Figure 13.13 compares modeled $CTF_{eye\text{-}disp,h}$ with measured $CTF_{eye\text{-}disp,h}$. In this section we develop *Mathematica* code that implements the modeled $CTF_{eye\text{-}disp,h}$.

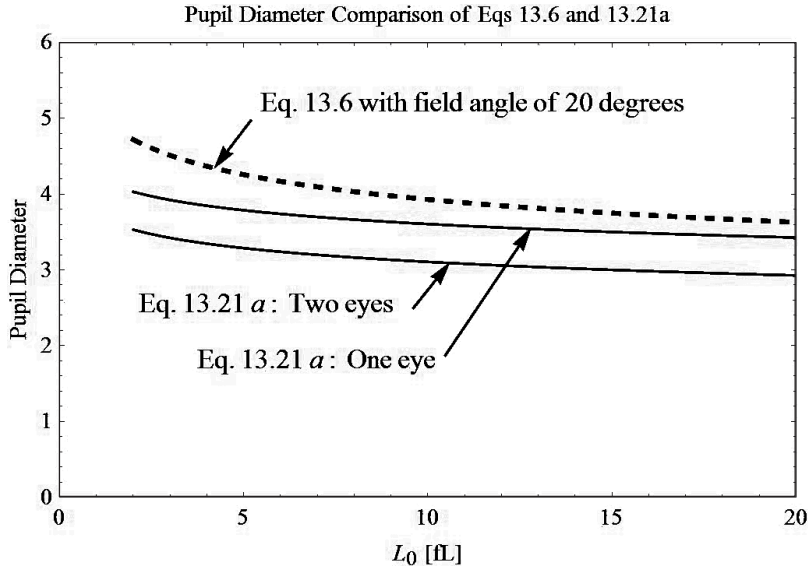


Figure 13.14 Comparison of pupil diameter as given by (13.6) and (13.21a).

Example 13.1
In (13.3) the frequencies in *CTF* were expressed in *cycle/deg* and the luminance values were expressed in *candela/m*^{2}. Reformulate (13.3) so ξ is expressed in *cycle/mr* and L_0 is expressed in *fL*. Then produce a graph similar to Figure 13.3 in which the horizontal axis is *cycle/mr* and the parametric curves are given in *fL*.

Solution: The conversion factor between *cycle/degree* and *cycle/mr* is obtained by direct calculation or from Wolfram Alpha (http://www.wolframalpha.com):

$$1\frac{\text{cycle}}{\text{deg}} = 1\frac{\text{cycle}}{\text{deg}}\frac{180\ \text{deg}}{\pi\ \text{radian}}\frac{1}{1000\dfrac{\text{mrad}}{\text{radian}}} = 0.05729\frac{\text{cycle}}{\text{mrad}}$$

$$\frac{1\ \text{cycle/deg}}{0.05729\ \text{cycle/mrad}} = 17.45\frac{\text{cycle/deg}}{\text{cycle/mrad}}$$

Wherever ξ occurs in (13.3) we replace it with 17.45 ξ where now the spatial frequency is in *cycle/mr*. Similarly, we find that in (13.3) we can replace L_0 with 3.425 L_0 providing L_0 is now expressed in *fL*:

$$CTF(\xi, L_0, w_h, \eta_{eye}) = min\left\{\frac{\sqrt{2}}{\sqrt{\eta_{eye}}}\left[a\,17.45\,\xi e^{-b17.45\xi}\sqrt{1+0.06e^{-b17.45\,\xi}}\right]^{-1}, 1\right\} \tag{13.36}$$

where
$$a = \frac{540\left(1+\dfrac{0.7}{3.425\ L_0}\right)^{-0.2}}{1+\dfrac{12}{w_h(1+0.333*17.45\xi)^2}}, \quad b = 0.3\left(1+\frac{100}{3.425 L_0}\right)^{0.15}$$

and $min(x_1,x_2)$ is the smaller of x_1,x_2. Equation (13.36) is illustrated in Figure 13.15. *Mathematica* code that implements the equations of Section 13.2 are presented next.

Eye–Display Contrast Threshold Model

The model takes the input parameters defined below in the *Define Input Parameters* section and outputs the eye–display contrast threshold function. Functions needed by the model are defined in the following *Define Functions* section. To calculate $CTF_{eye-disp,h}$, simply execute the command

```
CTFEyeDispH[ξ,L0,wh,nEye,N2DispH,σCR,σFP,CRT,γ,β,τDisp,X0]
```

and $CTF_{eye-disp,h}$ will be calculated for the given input parameters.

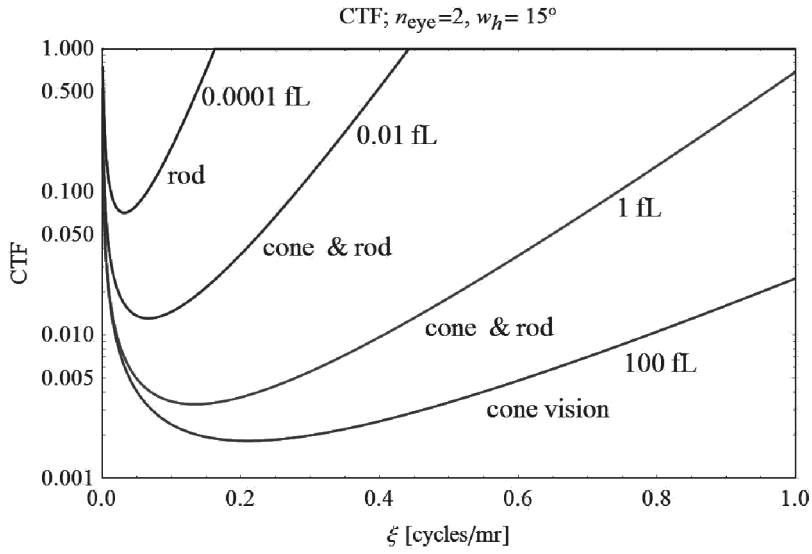


Figure 13.15 Contrast threshold function for two eyes as a function of spatial frequency for a variety of luminance values.

Define Functions

The *CTF* function below takes as input the average ambient luminance L_0 in *fL*, the width w_h of the sinusoidal display in Figure 13.1 in degrees, and the number of eyes used n_{Eye} and outputs the contrast threshold function as a function of the spatial frequency ξ in units of *cycle/mr* [per (13.36)]:

$$\begin{aligned}
&\text{CTF}[\xi_, L0_, wh_, nEye_] := \text{Module}\Big[\{a, b, temp\},\\
&\quad a = 540.\,\frac{\left(1+\frac{0.7}{3.425\,L0}\right)^{-0.2}}{1+\frac{12.}{wh\,(1+0.333\times 17.45\,\xi)^2}};\\
&\quad b = 0.3\left(1+\frac{100.}{3.425\,L0}\right)^{0.15};\\
&\quad temp = \frac{\sqrt{2}}{\sqrt{nEye}}\left(17.45\,a\,\xi\,e^{-17.45\,b\,\xi}\sqrt{1+0.06\,e^{-17.45\,b\,\xi}}\right)^{-1};\\
&\quad \text{Min}[temp, 1]\Big]
\end{aligned}$$

The time constant of the eye, τ_{eye} is a function of the average ambient luminance L_0 in *fL* [per (13.12)]:

$$\text{tEye}[L0_] := 4.\left(0.0192 + 0.0625\left(\frac{1.76}{L0}\right)^{0.17}\right)$$

The temporal noise bandwidth $Q_{Eye,t}$ in hertz is ([per (13.11)]:

$$\text{QEyet}[L0_] := \frac{1}{\text{tEye}[L0]}$$

The temporal noise bandwidth $Q_{eye-disp,t}$ is [per (13.17)]:

$$\text{QEyeDispt}[tDisp_, CRT_, L0_] := \begin{cases} \frac{1}{tDisp+tEye[L0]} & CRT \geq 0.5 \\ \frac{1}{tDisp}\left(1+\left(\text{Exp}\left[-\frac{2\,tDisp}{tEye[L0]}\right]-1\right)\frac{tEye[L0]}{2\,tDisp}\right) & CRT < 0.5 \end{cases}$$

We use the convention that if the calculation is for a *CRT* then *CRT* = 1; if the calculation is for a flat-panel display then *CRT* = 0. The display spatial transfer function H_{disp} is [per (13.19b) and (13.20b)]:

```
HDisp[ξ_, σCRT_, σFP_, CRT_] := { Exp[-π (σCRT ξ)^2]      CRT > 0.5
                                { Sin[π σFP ξ]/(π σFP ξ)   CRT < 0.5
```

The eye transfer function H_{eye} is also denoted by MTF_{eye} [per (13.21)]:

```
MTFEye[ξ_, L0_, nEye_] := Module[{dpupil, e0, fi, fe},
  dpupil = -9.011 + 13.23 Exp[-Log[10, L0]/21.082] - (nEye - 1) 0.5;
  e0 = (0.7155 + 0.277/Sqrt[dpupil])^2;
  fi = Exp[3.663 - 0.04974 dpupil^2 Log[10, dpupil]];
  fe = 43.63 ξ;
  (* Output the eye MTF *)
  Exp[-(fe/fi)^e0] Exp[-0.375 ξ^1.21] Exp[-0.4441 ξ]]
```

The channel *MTF*, H_{chan} is ([per (13.22)]):

```
HChan[ξp_, ξ_] := Exp[-2.2 (Log[Abs[ξp]/ξ])^2]
```

The eye–display bandwidth function $Q_{eye-disp,h}(\xi)$ is [per (13.18)]:

```
QEyeDispH[ξ_, σCRT_, σFP_, L0_, CRT_, nEye_] :=
 2 NIntegrate[(Abs[HDisp[ξp, σCRT, σFP, CRT] MTFEye[ξp, L0, nEye] HChan[ξp, ξ]])^2, {ξp, 0, ∞}]
```

We utilized the fact that the integrand is an even function of ξ' to convert (13.18) to a one-sided integral. The effective area of the eye pupil A_{eff} in square millimeters is [per (13.6) and (13.7)]:

```
Aeff[L0_, X0_] := Module[{dpupil},
  dpupil = 5 - 3 Tanh[0.4 Log[10, L0 3.426 X0^2/40^2]];
  (* output Aeff *)
  π/4 dpupil^2 (1 - (dpupil/9.7)^2 + (dpupil/12.4)^4)]
```

The variance internal to the eye σ^2_{eye} is [per (13.5b)]:

```
σ2Eye[L0_, X0_, nEye_, β_] := 1/nEye (1 + β^2 QEyet[L0]/(3.426 L0 Aeff[L0, X0]))
```

The conversion factor 3.426 in the code converts L_0 in *fL* to *candela/m*^{2}. The variance that describes the eye–display $\sigma^2_{eye-disp}$ is [per (13.5c)]:

```
σ2EyeDisp[N2DispH_, ξ_, σCRT_, σFP_, L0_, CRT_, nEye_, γ_, rDisp_, CRT_] :=
 L0^-2 γ^2 N2DispH QEyeDispH[ξ, σCRT, σFP, L0, CRT, nEye] QEyeDispt[rDisp, CRT, L0]
```

The contrast threshold function of eye–display system $CTF_{eye-disp,h}$ is [per (13.5a)]:

$$\text{CTFEyeDispH}[\xi_, L0_, wh_, nEye_, N2DispH_, \sigma CRT_, \sigma FP_, CRT_, \gamma_, \beta_, \tau Disp_, X0_] := \text{Min}\left[\text{CTF}[\xi, L0, wh, nEye]\sqrt{\left(1 + \frac{\sigma 2\text{EyeDisp}[N2DispH, \xi, \sigma CRT, \sigma FP, L0, CRT, nEye, \gamma, \tau Disp, CRT]}{\sigma 2\text{Eye}[L0, X0, nEye, \beta]}\right)}, 1\right]$$

Define Input Parameters

```
N2DispH = 4.5 ; (* Noise Power Sepectral Density at display [fl^2 mr s] *)

ξ = 0.1 ; (* Spatial frequency in horizontal direction [cycles/mr] *)

L0 = 8.4; (* Average display luminance [fL]   *)

wh = 0.94 ;(*  angle subtended at the eye of sinusoidal pattern [deg] *)


nEye = 2 ;(* number of eyes used; 1 or 2 *)

σCRT = 1.3 ;(* measure of CRT spot size; Eq 13.19; [mr]  *)

σFP = 1;(* measure of flat pannel spot size; Eq 13.20; [mr]  *)

CRT = 1; (* Value is 1 if CRT is used; Value is 0 for flat pannel display *)

γ = 330.;(* Calibration constant in Eq 13.5; Default value=330; [] *)

β = 4.9 ;(* Calibration constant in Eq. 13.5; Default value = 5; [√Trolands * sec ] *)

τDisp = 0.2(* Display time constant Eqs 13.14 and 13.15; [sec] *)

X0 = 20(* Horizontal angle subtended by display at eye; Eq 13.12; [degree] *)
```

Run The Program

Output intermediate and final results:

```
: CTF[ξ, L0, wh, nEye]

: 0.0123512

: τEye[L0]

: 0.268468

: QEyet[L0]

: 3.72484

: QEyeDispt[τDisp, CRT, L0]

: 2.13462

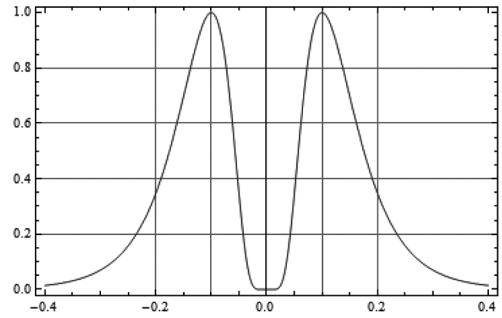
: HDisp[ξ, σCRT, σFP, CRT]

: 0.948292

: MTFEye[ξ, L0, nEye]

: 0.743774
```

```
Plot[HChan[ξp, ξ], {ξp, -0.4, 0.4}, GridLines → {{-.2, -.1, .1, .2}, Automatic}, Frame → True]
```




```
QEyeDispH[ξ, σCRT, σFP, L0, CRT, nEye]
0.0803029

Aeff[L0, X0]
10.6093

σ2Eye[L0, X0, nEye, β]
0.646458

σ2EyeDisp[N2DispH, ξ, σCRT, σFP, L0, CRT, nEye, γ, τDisp, CRT]
1190.51

CTFEyeDispH[ξ, L0, wh, nEye, N2DispH, σCRT, σFP, CRT, γ, β, τDisp, CRT]
0.563943
```

The computed value of $CTF_{eye\text{–}disp,h}$ is 0.56.

Comparison With Existing Models

Figure 13.16 compares calculations done here with the corresponding calculations done by the NV-IPM model for 20 sets of input parameters. The solid line corresponds to perfect NV-IPM agreement between the two calculations. Each point represents a pair of points calculated with *Mathematica* and with *NV-IPM* . The agreement between the two programs, which were developed independently, provides confidence that the model equations were correctly implemented.

13.5 TTP Metric and Range Performance Modeling

In this section we show how results obtained in Section 13.2 can be applied to the problem of predicting sensor range performance [15].

So far we have considered CTFs for which the sinusoidal variation is in the horizontal direction as shown in Figure 13.1. We want to extend the eye-display contrast threshold model to the case where the sinusoidal variation is in an arbitrary direction. We use ξ and η to denote spatial frequencies in the horizontal and vertical directions, respectively, and use ρ to indicate a spatial frequency in an arbitrary

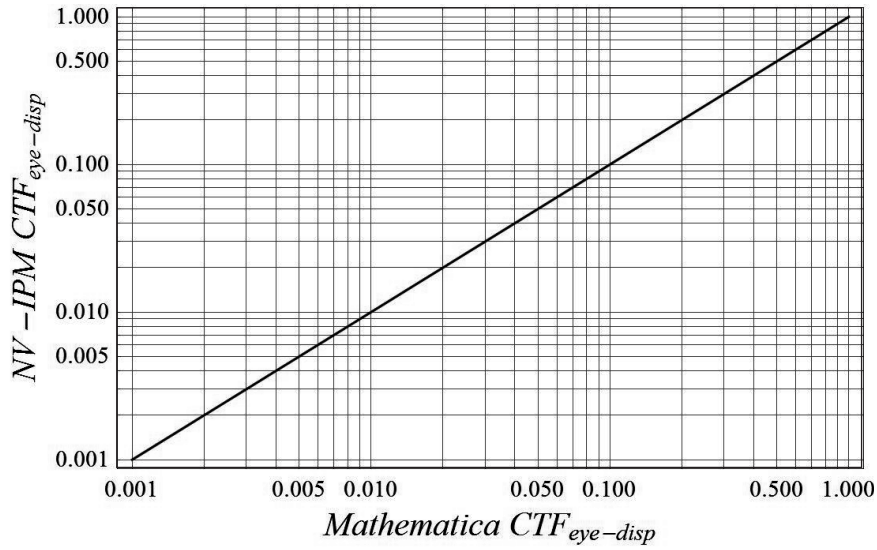


Figure 13.16 Comparison of $CTF_{eye\text{-}disp}$ using code described above and *NV-IPM* model.

direction. To a good approximation, we can assume that the contrast threshold function of the eye given by (13.3) is the same in the horizontal and vertical directions. The eye–display contrast threshold in the vertical direction can be different from that in the horizontal direction because sensor and display design may favor one direction over another but typically imaging sensors are designed so that the performances in the horizontal and vertical directions are comparable.

We now define the spatial frequency ρ in an arbitrary direction and exhibit the relationship between ρ, ξ, and η. Figure 13.17 shows a sinusoidal intensity variation tilted at an angle θ. Here the dark lines correspond to the lines of minimum illuminance in Figure 13.1. The period of the sinusoidal variation is L in the direction θ and the period is L_x and L_y in the x and y directions, respectively. For the illustrated sinusoidal pattern

$$\rho \equiv 1/L; \quad \xi \equiv 1/L_x; \quad \eta \equiv 1/L_y \tag{13.37a}$$

This is a generalization of the equation for ξ illustrated in Figure 13.2. From Figure 13.17, it is apparent that when $\theta = 0$, $\rho = \xi$ and $\eta = 0$. It is also apparent that when $\theta = \pi/2$, $\rho = \eta$, and $\xi = 0$. These are special cases of a more general relationship between ρ, ξ, and η:

$$\rho = \sqrt{\xi^2 + \eta^2} \tag{13.37b}$$

Because targets have extent in the horizontal and vertical directions, we need an expression for the *system contrast threshold function* in an arbitrary direction. Let $CTF_{eye\text{-}disp,h}(\xi)$, $CTF_{eye\text{-}disp,v}(\eta)$, $H_{sys,h}(\xi)$, and $H_{sys,v}(\eta)$ denote the system contrast threshold functions and system transfer functions in the horizontal and vertical directions, respectively. The *system transfer function* in the horizontal direction $H_{sys,h}(\xi)$ is the ratio of the target modulation to the display modulation in the horizontal direction. A similar definition applies to $H_{sys,v}(\eta)$ The system contrast threshold functions in the horizontal and vertical directions are

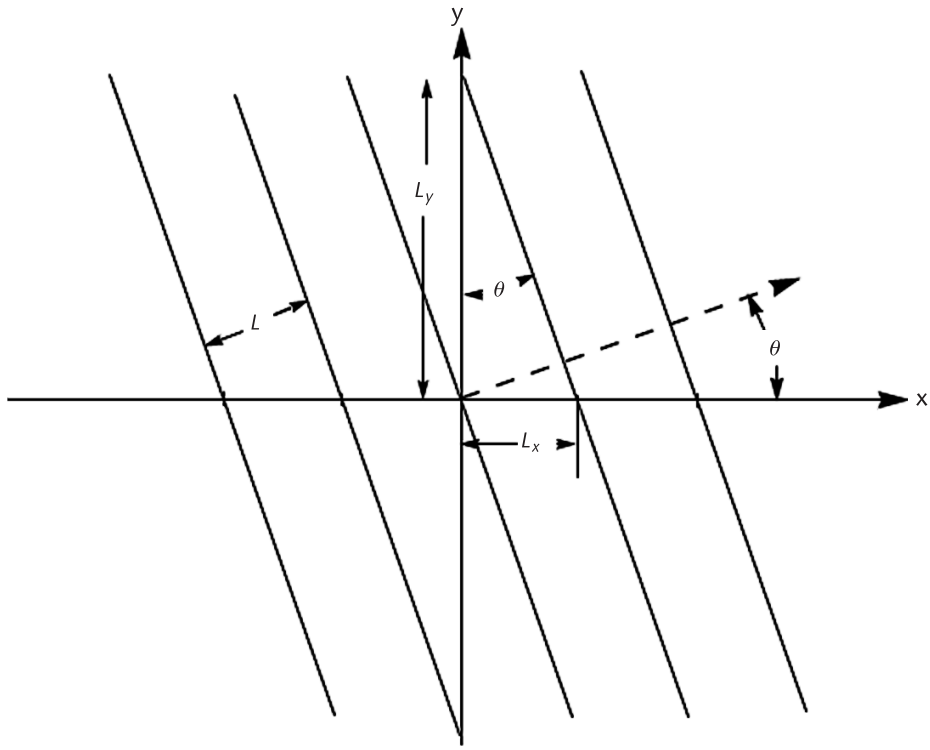


Figure 13.17 Spatial periods of a sinusoidal variation tilted at an angle.

$$CTF_{sys,h}(\xi) = \frac{CTF_{eye-disp,h}(\xi)}{H_{sys,v}(\eta)} \tag{13.38a}$$

$$CTF_{sys,v}(\eta) = \frac{CTF_{eye-disp,v}(\eta)}{H_{sys,v}(\eta)} \tag{13.38b}$$

These equations assert that the system contrast threshold function goes up as the system transfer function goes down, which is in agreement with both intuition and experimental observations. Typically $CTF_{sys,h}(\xi)$ and $CTF_{sys,v}(\eta)$ can be calculated from the sensor design using the results of Section 13.2 and Chapter 14.

In a typical system $CTF_{sys,h}(\xi)$ and $CTF_{sys,v}(\eta)$ are comparable at a given frequency, and we estimate the CTF_{sys} in an arbitrary direction as the geometric mean of the contrast threshold functions in the horizontal and vertical directions:

$$CTF_{sys}(\rho) \cong \sqrt{CTF_{sys,h}(\rho)CTF_{sys,v}(\rho)} \tag{13.38c}$$

We assume that $CTF_{sys}(\rho)$ is approximately independent of θ.

In agreement with intuition, the ability to detect, recognize, or identify a target depends on target size and contrast.

The target size is characterized by the angle θ_{tgt} subtended at the sensor by the target. Let A_{tgt} denote the *projected area* of the target. The *projected area* of the target is defined as the area of the shadow formed by the target on an imaginary

screen perpendicular to the sensor's line of sight if there were a point source at the sensor. A *target characteristic dimension* L_{tgt} is then defined by

$$L_{tgt} \equiv \sqrt{A_{tgt}} \quad [\text{m}] \tag{13.39a}$$

Typically the target area is measured in m^2 so the dimension of L_{tgt} is m. An average angle θ_{tgt} subtended by the target at the sensor is then given by

$$\theta_{tgt} \equiv \frac{L_{tgt}}{R} \quad [\text{mrad}] \tag{13.39b}$$

where R is the sensor-target range. Typically the range is measured in *km* and then with L_{tgt} measured in m, θ_{tgt} given by Eq (13.39b) is in *mr*.

The *target contrast* C_{tgt} is calculated using (13.2a) where $L_{\max}$ is a measure of the maximum target luminance on the display, and $L_{\min}$ (is a measure of the minumum background luminance of the display. Here we are assuming the target, as shown on the display, is brighter than the background. If the target is dimmer than the background, then $L_{\max}$ is a measure of the maximum background luminance and $L_{\min}$ is a measure of the minimum target luminance.

We seek a *target transfer performance* metric that combines the target contrast with the system contrast threshold function $CTF_{sys}(\rho)$. The intent of the *TTP* metric is to describe the ease of doing a detection, recognition, or identification task. As expected, it gets easier to do any of these tasks as the target contrast goes up and it gets easier to do any of the tasks as $CTF_{sys}(\rho)$ goes down. A *TTP* metric that satisfies these conditions is

$$TTP \equiv \left(\frac{C_{tgt}}{CTF_{sys}(\rho)} \right)^{\alpha}$$

where $\alpha > 0$. A problem with this definition is that it does not specify what spatial frequency ρ to use. It seems reasonable to use all spatial frequencies for which $C_{tgt} > CTF_{sys}(\rho)$. Thus a better guess at the *TTP* metric is

$$TTP \equiv \int_{ut\text{-}on}^{cut\text{-}off} \left(\frac{C_{tgt}}{CTF_{sys}(\rho)} \right)^{\alpha} d\rho \quad [\text{cycle/mrad}] \quad \text{(second guess)} \tag{13.40a}$$

where *cut-on* and *cut-off* are the spatial frequencies ρ where $C_{tgt} = CTF_{eye-disp}(\rho)$ Since contrasts are dimensionless and ρ has the dimensions cycle/mrad, the TTP metric has dimensions *cycle/mrad*.

We now seek a function V that describes how easy it is to find a target. The larger V is, the easier it is to find the target. Because it gets easier to find the target as θ_{tgt} increases and it gets easier to find the target as *TTP* increases, it is reasonable to define V by the relationship

$$V \equiv TTP \cdot \theta_{tgt} = \left(\int_{cut\text{-}on}^{cut\text{-}off} \left(\frac{C_{tgt}}{CTF_{sys}(\rho)} \right)^{\alpha} d\rho \right) \frac{\sqrt{A_{tgt}}}{R} \quad [\text{cycles}] \tag{13.41a}$$

For a variety of targets under a variety of conditions we calculate for several values of. Experimentally, it is found that $\alpha = 1/2$ most consistently describes the task difficulty (be it detection, recognition, or identification). Thus, (13.40a) becomes

$$TTP \equiv \int_{cut\text{-}on}^{cut\text{-}off} \sqrt{\frac{C_{tgt}}{CTF_{sys}(\rho)}}\, d\rho \quad \text{[cycle/mrad]} \tag{13.40b}$$

and (13.41a) becomes

$$V(R) \equiv \left(\int_{cut\text{-}on}^{cut\text{-}offf} \sqrt{\frac{C_{tgt}}{CTF_{sys}(\rho)}}\, dp \right) \frac{\sqrt{A_{tgt}}}{R} \quad \text{[cycles]} \tag{13.41b}$$

There is an approximation is often used to simplify (13.40b). As already pointed out, the *cut-on* and *cut-off* frequencies ρ are the low and high spatial frequencies for which $C_{tgt} = CTF_{sys}(\rho)$. Figure 13.3 shows the shape of the contrast threshold function for the eye looking at a display without noise and with $\tau_{disp} >> \tau_{eye}$. The effect of using a noisy display is for the *CTF* to keep roughly the same shape, but the curve moves upward. If the target contrast is less than the smallest value of the contrast threshold function in Figure 13.3 then there is no solution for the *cut-on* and *cut-off* frequencies, implying that V is 0 and the target cannot be seen. More typically, the target contrast is substantially greater than *CTF* and we see from Figure 13.3 that the *cut-on* frequency will be approximately 0. Thus,

$$TTP \cong \int_{0}^{cut\text{-}off} \sqrt{\frac{C_{tgt}}{CTF_{sys}(\rho)}}\, d\rho \quad \text{[cycle/mrad]} \tag{13.40c}$$

Because of the way V was constructed, the probability of successfully doing a task depends on V: When V is small, the probability of successfully doing the task is small; as V increases, the probability of successfully doing the task increases. Thus, we seek a probability as a function of range $P(R)$ with the property that $P(R) = 0$ when $V(R) = 0$ and which monotonically increases to 1 when $V(R)$ gets sufficiently large. A function with these properties is

$$P(R) = \frac{\left(\frac{V(R)}{V_{50}} \right)^{\alpha}}{1 + \left(\frac{V(R)}{V_{50}} \right)^{\alpha}} \quad \text{(first guess)} \tag{13.42a}$$

where α is a constant to be determined. Figure 13.18 illustrates this function for several values of α.

The value of α is determined using the following procedure: Decide on a task (typical tasks are detect, recognize, or identify a target), decide on a target or set of targets, and then calculate $V(R)$ as a function of range using (13.41b). Vary the range until the probability for doing the task is 0.5. The value of V at this range is V_{50}, which is now known. With the known value of V_{50}, the probability of doing the task can be experimentally measured as a function of $V(R)$. Using a least-squares procedure, this enables α to be found for a particular target set, for a particular sensor, and for a particular

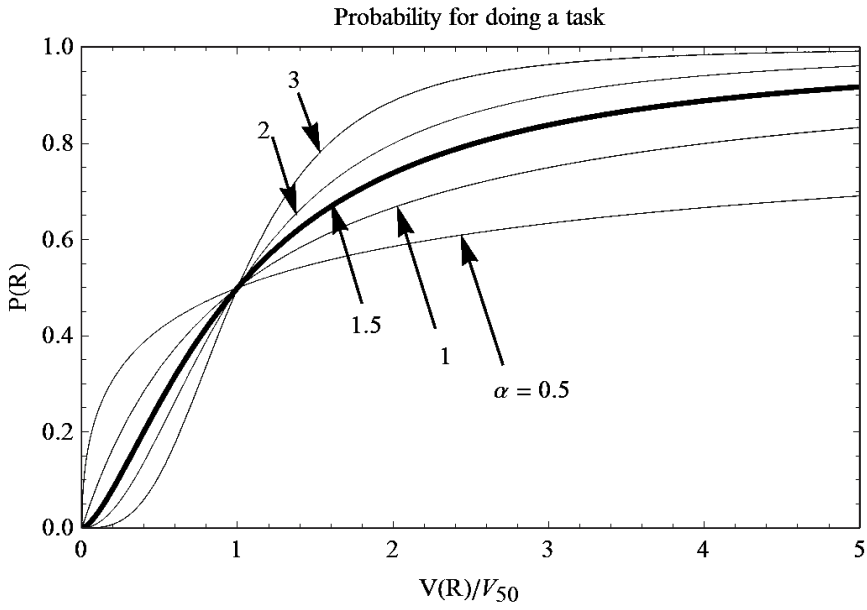


Figure 13.18 Illustrating (13.42a) for several values of á.

background. The experiment is done many times for different target sets, sensors, and backgrounds. We find that V_{50} depends strongly on the task but under a wide variety of conditions α is approximately constant and has the value 1.5. Thus, (13.42a) becomes

$$P(R) = \frac{\left(\frac{V(R)}{V_{50}}\right)^{1.5}}{1+\left(\frac{V(R)}{V_{50}}\right)^{1.5}} \tag{13.42b}$$

The thick curve in Figure 13.18 corresponds to and shows how $P(R)$ grows as the ratio $V(R)/V_{50}$ grows.

13.6 Guide to the References

The work [9, 10] by Preece et al. is among our favorite references for the application of CTF for predicting sensor performance. Barton has a succinct [1] discussion of CTF and a comprehensive [2] treatment. Szirtes [13] has a comprehensive dimensional analysis treatment with many examples and is arguably the best book on the subject. The treatment by Bernstein [3] et al. is a good introductory treatment.

13.7 Exercises

13.1 Eye doctors find out how well your eyes are working using a Snellen eye chart. Why is it necessary to use the contrast threshold function rather than something akin to the Snellen eye chart to predict sensor performance?

13.2 B. G. Grant in *Field Guide to Radiometry*, SPIE Press, p. 91 (2011) gives the following information: Scotopic and photopic vision corresponds to luminance values of less than 0.003 and greater than 3 cd/m^2 respectively. What luminance values do these correspond to in fL?

13.3 The dimensions of $N^2_{disp,h,t}$ in (13.5a) are fL^2 *mrad·s*. This noise power spectral density describes a CTF in one spatial dimension as exhibited by the observation that $Q_{eye-disp,h}(\xi)$ is the only spatial factor multiplying $N^2_{disp,h,t}$. If we had a noise power spectral density that described two spatial directions so that it also included $Q_{eye-disp,v}(\eta)$ then what would the dimensions of the noise power spectral density be?

13.4 The average luminance on a display is 1.0 fL and the standard deviation of the display luminance is 0.1 fL The display is 640 × 480 pixels, is 25 × 18.7 cm, and is viewed from a distance of 60 cm. The frame rate is 60 *frames/s*. What is the noise power spectral density in the horizontal and vertical directions?

13.5 Produce a graph of the contrast threshold function of the naked eye assuming the average display luminance is 15 fL.

13.6 For a display with L_0 = 1.0 fL, calculate the *cut-on* and *cut-off* frequencies for a target that has a display contrast of 0.5. Calculate the *TTP* metric using the *cut-on* and *cut-off* frequencies and compare with the *TTP* metric assuming the *cut-on* frequency is zero. The purpose of this problem is to investigate the validity of the approximation in going from (13.40b) to (13.40c).

13.7 Using the parameters of Section 13.4.3, but with $N^2_{disp,h,t}$ equal to 0.0, 0.1, 0.5, 1 and 5, calculate $CTF_{eye-disp,h}$ as a function of ξ.

13.8 A target has a characteristic dimension of 1.5m at a range of 1 km. The target contrast on the display is 0.3, the average display luminance is 8.4 fL, and $N^2_{disp,h,t}$ can be approximated as 0. Calculate the *TTP* metric and *V*. If V_{50} = 10.0 for recognition, estimate the probability for recognizing the target at a range of 1 km.

13.9 The definition for CTF_{sys} given by (13.38c) is the geometric mean. Alternate definitions for CTF_{sys} could be based on the arithmetic or harmonic mean:

$$CTF_{sys,arith}(\rho) \equiv 0.5(CTF_{sys,h}(\rho) + CTF_{sys,v}(\rho)) \tag{13.43}$$

$$CTF_{sys,har}(\rho) \equiv \frac{2CTF_{sys,h}(\rho)\, CTF_{sys,v}(\rho)}{CTF_{sys,h}(\rho) + CTF_{sys,v}(\rho)} \tag{13.44}$$

Let *A*, *H*, and *G* denote the arithmetic, harmonic, and geometric means, respectively. Recall the result from http://mathworld.wolfram.com/ArithmeticMean.html that $A \geq G \geq H$ so the geometric mean is always less than or equal to the arithmetic mean and is always greater than or equal to the harmonic mean. Typically $CTF_{sys}(\rho)$ decreases monotonically with increasing ρ regardless of whether it is based on the arithmetic, harmonic, or geometric mean. This implies the *TTP* metric and $V(R)$ will decrease with range. If the CTF_{sys} definition were changed, do you predict that the only thing that would need to be changed to model sensor performance is either to change the form of

(13.42a) or use a different value for α in that equation? Do you agree that there is some arbitrariness in how CTF_{sys} is related to the horizontal and vertical contrast threshold functions, but this arbitrariness can be compensated for by properly choosing the empirical relationship relating $P(R)$ and $V(R)$?

13.10 As indicated in Figure 13.1, (13.3) is one dimensional in the sense that there is dependence on w_h but no dependence on w_v. We believe (13.3) is a good approximation providing $w_v > 15°$ and that it is desirable to modify this equation so that it applies when deg. Because the bars are in the vertical direction, it is possible that w_h and w_v enter the equation asymmetrically. However, let us assume they enter the equation symmetrically. Then a proposed generalization of (13.3) is

$$CTF(\xi, L_0, w_h, w_v, n_{eye}) = min\left\{\frac{\sqrt{2}}{n_{eye}}\left[a\,\xi\, e^{-b\xi}\sqrt{1+0.06e^{-b\xi}}\right]^{-1}, 1\right\} \qquad (13.45)$$

where

$$b = 0.3\left(1+\frac{100}{L_0}\right)^{0.15}$$

$$a = \begin{cases} \dfrac{540\left(1+\dfrac{0.7}{L_0}\right)^{-0.2}}{1+\dfrac{12}{w_h(1+0.333\xi)^2}} & \text{for } w_v \geq 15°, w_h > 0.3° \\ \dfrac{540\left(1+\dfrac{0.7}{L_0}\right)^{-0.2}}{1+\dfrac{12}{w_h(1+0.333\xi)^2}} & \text{for } w_v \leq 15°, w_h > 0.3° \end{cases}$$

An alternate approach is to assume (13.3) is valid but replace w_h with $\sqrt{w_h\, w_v}$. Graph (13.45) and the modified form of (13.3) as a function of w_v to see how different the two expressions are. Because w_h and w_v for targets on the display are typically much less than 15°, this indicates the importance of doing experiments to improve (13.3) so that it accurately predicts a CTF appropriate for tactical targets.

Appendix 13A: Direct Calculations of $CTF_{eye-disp,h}$

In this appendix, (13.4) and $CTF_{eye-disp,h}$ (13.5a) are calculated directly from the properties of the eye without invoking dimensional analysis techniques. The derivation given here follows that given by Preece[9] et al.

Derivation of Equation (13.4)

Let σ^2_{eye} denote the total internal eye–brain noise variance, S_{det} the perceived signal in the visual cortex, and Ψ the threshold such that the signal is detected 50% of the

time. We hypothesize that σ_{eye}^2 depends on the ambient luminance L_0 and the number of eyes n_{eye}, and that S_{det} depends on the spatial frequency ξ of the displayed pattern, the width w_h shown in Figure 13.1, and on the ambient luminance L_0. We assume the threshold function Ψ is the ratio of the detected signal to the internal eye–noise standard deviation:

$$\Psi\left(\xi, L_0, w_h, n_{eye}\right) = \frac{S_{det}\left(\xi, L_0, w_h\right)}{\sigma_{eye}\left(L_0, n_{eye}\right)} \tag{13A.1}$$

A first guess is that

$$S_{det} = c_1 L_p A_{eff} H_{eye}(\xi) H_{lat}(\xi) \phi\left(w_h\right) \quad (\textit{first guess}) \tag{13A.2a}$$

where c_1 is a dimensionless constant, L_p is the luminance amplitude illustrated in Figure 13.2, A_{eff} is the effective pupil area of the eye, and H_{eye} and H_{lat} are the MTF of the eye including the lens, tremor, and retina bandpass filtering, while H_{lat} is the MTF from lateral inhbition. The signal is expected to grow with increasing w_h and this is described by the function $\phi(w_h)$. A problem with (13A.2a) is that it does not reflect Weber's [5, 15] law: The threshold needed is proportional to the mean luminance. A guess for S_{det} that recognizes Weber's law and makes S_{det} dimensionless is as follows:

$$S_{det} = c_1 \frac{L_p A_{eff} H_{eye}(\xi) H_{lat}(\xi) \phi\left(w_h\right)}{L_0\, A_{eff}} \tag{13A.2b}$$

The signal is assumed to be static, which is why there is no temporal *MTF* in (13A.2b). Substituting into (13A.1), we get

$$\Psi\left(\xi, L_0, w_h, n_{eye}\right) = c_1 \frac{L_p}{L_0} \frac{H_{eye}(\xi) H_{lat}(\xi)\, \phi\left(w_h\right)}{\sigma_{eye}\left(L_0, n_{eye}\right)} \tag{13A.3}$$

Solve (13A.3) for the ratio L_p/L_0 and realize from (13.2b) that this is the contrast. Thus, the contrast threshold function of the naked eye is

$$CTF = \frac{1}{c_1} \frac{\Psi\left(\xi, L_0, w_h, n_{eye}\right)\ \sigma_{eye}(L_0, n_{eye})}{H_{eye}(\xi) H_{lat}(\xi) \phi\left(w_h\right)} \tag{13A.4}$$

We seek a generalization of (13A.4) that is valid when the eye is looking at a noisy display. Let $\sigma_{eye-disp}^2$ denote the perceived external noise variance of the eye–brain system and let $CTF_{eye-disp,h}$ denote the con trast threshold function of the eye–display system viewing a noisy display. We assume that the eye–display variance is the sum of σ_{eye}^2 and $\sigma_{eye-disp}^2$. Thus the generalization of (13A.4) is

$$CTF_{eye-disp,h} = \frac{1}{c_1} \frac{\Psi\left(\xi, L_0, w_h, n_{eye}\right) \sqrt{\sigma_{eye}^2\left(L_0, n_{eye}\right) + \sigma_{eye-disp}^2\left(\xi, L_0\right)}}{H_{eye}(\xi) H_{lat}(\xi) \phi(w_h)} \tag{13A.5}$$

In writing (13A.5) from (13A.4), we assumed that the threshold function Ψ is independent of the noise on the display. Squaring (13A.5) and (13A.4) and

forming the ratio, we find that the constant c_1 and the functions Ψ, H_{eye}, and H_{lat} cancel:

$$CTF^2_{eye-disp,h} = CTF^2\left(1+\frac{\sigma^2_{eye-disp}\left(\xi, L_0\right)}{\sigma^2_{eye}\left(L_0, n_{eye}\right)}\right) \tag{13A.6}$$

This is in agreement with (13.4), one of the fundamental results of Section 13.2.

Derivation of Equation (13.5a)

The quantity $\sigma^2_{eye-disp}$ is proportional to the noise power spectral density $N^2_{disp,h,t}$, which has units $fL^2 \cdot mrad \cdot s$. Thus,

$$\sigma^2_{eye-disp} = c_2 \frac{N^2_{disp,h,t}\, Q_{eye\text{-}disp,h}(\xi)\, Q_{eye\text{-}disp}}{L_0^2} \tag{13A.7}$$

where c_2 is a dimensionless constant and $Q_{eye-disp,h}$ and $Q_{eye-disp,t}$ have units of mrad^{-1} and sec^{-1} respectively. The quantity L_0^2 in the denominator has units of fL^2 consistent with the view that $\sigma^2_{eye-disp}$ is dimensionless.

The quantity σ^2_{eye} is proportional to the sum of a photon variance σ^2_{photon} and a floor variance σ^2_{floor}:

$$\sigma^2_{eye} = \left(\sigma^2_{floor} + \sigma^2_{photon}\right) \text{ (first guess)} \tag{13A.8a}$$

The floor variance is needed to ensure that σ^2_{eye} does not get arbitrarily small for arbitrarily small values of σ^2_{photon}. This ensures that we do not divide by a small number in (13A.6), is supported by CTF measurements made on noisy displays, and is consistent with the observation that the minimum contrast threshold function of the eye never goes to 0 (see Figure 13.3). A first guess is that

$$\sigma^2_{photon} \;\mu L_0\, A_{eff}\, Q_{eye,t} \;\text{ (first quess)} \tag{13A.9a}$$

where L_0, A_{eff}, and $Q_{eye,t}$ are the ambient luminance, effective pupil area, and temporal bandwidth of the eye. Equation (13A.9a) reasonably asserts that the photon variance goes up with the temporal bandwidth of the eye. Consistent with a Poisson probability density function for the number of photons, (13A.9a) asserts that σ^2_{photon} is proportional to $L_0\, A_{eff}$. However this guess is not consistent with contrast threshold function measurements made in the presence of noise. Furthermore (13A.9a) is not dimensionless, which is inconsistent with (13A.8a). To get results consistent with CTF measurements, we introduce a dimensional constant κ with units of $\sqrt{Troland \cdot s}$. Then the relationship (13A.9a) becomes

$$\sigma^2_{photon} = \kappa^2 \frac{L_0\, A_{eff}\, Q_{eye,t}}{(L_0\, A_{eff})^2} \tag{13A.9b}$$

and we note that σ^2_{photon} is now dimensionless. Then (13A.8a) becomes

$$\sigma^2_{eye} = \left(\sigma^2_{floor} + \kappa^2 \frac{Q_{eye,t}}{L_0 A_{eff}} \right) \text{ (first quess)} \tag{13A.8b}$$

In (13A.8b) we have not considered the number of eyes. Because the noise sources in the eye are independent, it is expected that σ^2_{eye} will be inversely proportional to the number of eyes used. Thus (13A.8b) becomes

$$\begin{aligned} \sigma^2_{eye} &= \frac{1}{n_{eye}} \left(\sigma^2_{floor} + \kappa^2 \frac{Q_{eye,t}}{L_0 A_{eff}} \right) \\ &= \frac{\sigma^2_{floor}}{n_{eye}} \left(1 + \frac{\kappa^2}{\sigma^2_{floor}} \frac{Q_{eye,t}}{L_0 A_{eff}} \right) \end{aligned} \tag{13A.8c}$$

Substituting (13A.7) and (13A.8c) into (13A.6), we obtain

$$CTF^2_{eye-disp,h} = CTF^2 \left(1 + \frac{C_2 \dfrac{N^2_{disp,h,t}\, Q_{eye-disp,h}(\xi) Q_{eye-disp,t}}{L_0^2}}{\dfrac{\sigma^2_{floor}}{n_{eye}} \left(1 + \dfrac{\kappa^2}{\sigma^2_{floor}} \dfrac{Q_{eye,t}}{L_0 A_{eff}} \right)} \right)$$

$$= CTF^2 \left(1 + \frac{\gamma_2 \dfrac{N^2_{disp,h,t}\, Q_{eye-disp,h}(\xi) Q_{eye-disp,t}}{L_0^2}}{\dfrac{1}{n_{eye}} \left(1 + \beta^2 \dfrac{Q_{eye,t}}{L_0 A_{eff}} \right)} \right) \tag{13A.9c}$$

where

$$\gamma^2 \equiv \frac{C^2}{\sigma^2_{floor}}; \quad \beta^2 \equiv \frac{\kappa^2}{\sigma^2_{floor}}$$

Equation (13A.9c) is in agreement with (13.5a).

References

[1] Barton, P. G. J., "Evaluation of the effect of noise on subjective image quality," *Human Visual Processing and Digital Display II, SPIE*, Vol. 1453, 1991.

[2] Barton, P. G. J., "Contrast Sensitivity of the Human Eye and Its Effects on Image Quality," *SPIE – The International Society for Optical Engineering*, 1999, p. 31.

[3] Bernstein, M. A., and W. A. Friedman, *Thinking about Equations: A Practical Guide for Developing Mathematical Intuition in the Physical Sciences and Engineering*, New York City, NY: John Wiley & Sons, 2009.

[4] Bridgman, P. W., *Dimensional Analysis*, Forgotten Press, 2010.

[5] Daniels, A., *Field Guide to Infrared Systems, Detectors, and FPAs, Second Edition*, Bellingham, WA: SPIE Press, 2010.

[6] Langhaar, H. L., *Dimensional Analysis and Theory of Models*, New York City, NY: John Wiley & Sons, 1951.

[7] Overington, I., *Vision and Acquisition*, London, UK: Pentech Press, 1976.

[8] Preece, B. L., J. T. Olson, J. P. Reynolds, and J. D. Fanning, "Improved noise model for the US Army sensor performance metric," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, G. Holst and K. Krapels (eds.), *Proc. of SPIE* , Vol. 8014.

[9] Preece, B. L., J. T. Olson, J. P. Reynolds, and J. D. Fanning, "Human vision noise model validation for the US Army sensor performance metric", *Proc. of SPIE*, Vol. 8014, 2011.

[10] Schwiegerling, J., *Field Guide to Visual and Opthalmic Optics*, Bellingham, WA: SPIE Press, 2004, p. 66.

[11] Stiles, W. S., and B. H. Crawford, "The luminous efficiency of the rays entering the eye pupil at different points," *Proc. Roy. Soc.*, London, UK, 1933, pp. 428–450.

[12] Szirtes, T., *Applied Dimensional Analysis and Modeling*, Oxford, UK: McGraw-Hill, 1998.

[13] Yoder, P. R. Jr., and D. Vukobratovich, *Field Guide to Binoculars and Scopes*, Bellingham, WA: SPIE Press, 2011, p. 8.

[14] Vollmerhausen, R. H., E. Jacobs, J. Hixson, and M. Friedman, "The Targeting Task Performance (TTP) metric: A new model for predicting target acquisition performance," *Technical Report AMSEL-NV-TR-230*, 2006.

[15] Vollmerhausen, R. H., E. Jacobs, and R. G. Driggers, "New metric for predicting target acquisition performance," *Opt. Eng*, Vol. 43, No. 11, 2004, pp. 2806–2824.

[16] http://en.wikipedia.org/wiki/Troland, last accessed May 2011.

CHAPTER 14

Infrared and EO System Performance and Target Acquisition

In this chapter, we describe how overall imaging system performance is calculated. It includes all of the sensitivity and resolution characteristics of the components covered in the past chapters. We begin this chapter in the infrared, but all of the concepts are applicable to electro-optical (reflective) systems.

Infrared systems underwent initial development just before World War II [1]. Due to the success of radar, imaging I^2R systems were investigated primarily in the laboratory until the 1960s. Once developed, their highest value was identified as providing battlefield observation in the dark. Infrared sensors allowed both targeting and fire control with complete day and night coverage. Huge successes were seen in Desert Storm with the night superiority provided by I^2R systems. The techniques presented in this chapter are the same as those used for the design and analysis of these military systems.

The FLIR receiver is a military version of the I^2R sensor. Both terms, FLIR and I^2R, are used interchangeably in the tactical and strategic sensor communities. The progress in infrared imaging technology has been described in terms of the FLIR. The first-generation FLIR is widely fielded on many platforms including tanks, helicopters, and rifles (also, handheld sighting systems). The first-generation FLIR is generally called a *common-module FLIR*. This stems from the use of system components made up of modules designed to be configured many different ways to be compatible with the platform constraints. The modules or building blocks allow system versatility in design while minimizing system uniqueness and, therefore, major expense throughout the life cycle of the equipment. Minimized logistic costs are realized because many of the components can be interchanged with similar fielded common-module systems. Due to specific system weight, power, and size constraints, however, some of the common modules are unique and cannot be interchanged. For example, there is a large IR imager module and a small IR imager module. The first-generation FLIR may comprise a linear array of detectors that are optically swept across the sensor FOV. Three detector array modules consist of 60, 120, and 180 detectors each. The 120- and 180-element modules are interchangeable, whereas the 60-element module is unique to itself and is used for rifle and handheld applications. The 120-element module is typically used for ground vehicle platforms, whereas the 180-element module is used for airborne applications.

There are two unique and noninterchangeable scanner modules: one for the small 60-element detector module and the other for the 120- and 180-element detector modules. The scanners are designed in such a way that the scanning mirror tilts on successive scans, thus providing successive interlace fields to the infrared imagery. The scanners are bidirectional in operation. All of the electronic amplifiers and filters are

analog. The infrared-to-visible light conversion is performed with an EO mux. The first-generation FLIR was very successful in Desert Storm. First-generation FLIRs have been fielded in Apache helicopters, light observation helicopters, M1A tanks, Bradley fighting vehicles, F-16 fighters, and with troops (the AN/TAS-4A), and the Night Observation Device–Long Range (NODLR), to name a few.

Second-generation FLIRs have been under development and deployment for the past 20 years. They are well defined and have been patented [Blecha, et. al., U.S. Patent Number 5,510,618]. The systems were intended to replace the fielded first-generation systems initially, and then become integrated components in new weapon system platforms. The major differences are that the detector has the ability to perform TDI processing with a number of detectors in a single scan line. Typically, the detector is 480×4, where the four-detector width is scanned and read out serially to perform the TDI function. In short, each of the four detectors samples the FOV at the same point and the signal is integrated to increase its fidelity. The signal adds directly and the noise adds in quadrature giving an increased SNR by a factor of 2. On-chip electronics are still analog, but digital processing is performed off the chip. Second-generation FLIRs are still single-band systems that either operate in the midwave or longwave spectral bands.

The third-generation FLIR has been under development for the past 10 years and has recently been fielded on a limited number of platforms. The third-generation FLIR technologies are listed in Figure 14.1. Third-generation FLIRs include a large-format staring array that has two bands (midwave and longwave).

Lloyd [2] points out that the imaging process has several different dimensions: intensity, time, wavelength, polarization, and the three spatial dimensions. With the

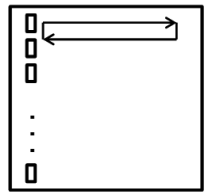
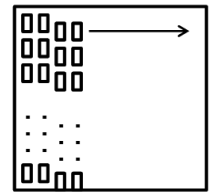
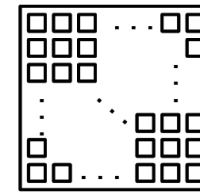
First Generation	Second Generation	Third Generation
		
Bi-Directional Linear Array	Uni-Directional TDIx4 Array	Large Format Array
Analog Electronics	Analog On-chip Digital Off-chip	Analog/Digital On-Chip
EO-Mux/E-Mux Analog Filtering	Digital Signal Processing	Digital Signal Processing
Single Waveband	Single Waveband	Dual Band (MWIR and LWIR)

Figure 14.1 FLIR generations.

number of variables contained in these dimensions, there are many correct solutions for a given imaging problem. Unfortunately, a large number of incorrect parameter sets can also be taken as solutions to a particular situation. Analysis and design go hand in hand; they form an iterative process. We assume that the analysis lends itself to linear shift-invariant principles in the development of the response for an infrared sensor design.

Performance measures are an excellent method for representing the collective response of a sensor system. However, infrared sensor performance measures do not include all of the necessary information to determine how well a sensor will perform in an overall imaging environment (e.g., a tactical or strategic scenario). That is, the I^2R system performance measures do not explicitly include target, background, or atmospheric characteristics. The measures only include the sensor components as outlined by the dashed box in Figure 14.2. Although the sensor system does not include the target, background, and atmospheric components, it does include the display and human vision components (if they are intended to process the data).

The idea here is to characterize the sensor with a number of system-level performance parameters that describe the fidelity of the infrared-to-visual transformation. These system-level metrics include the response of each sensor component. They can be used to determine how well a sensor performs in an overall imaging system scenario for particular targets, backgrounds, and atmospheric conditions, thus allowing a quick evaluation of the sensor in different environments.

The system-level performance measures describe the sensor in terms of sensitivity and resolution. Noise equivalent temperature difference (NETD) and 3D noise parameters describe the sensor's general sensitivity. The system modulation transfer function (MTF) describes the imaging resolution of the sensor. However, sensitivity and resolution of a sensor are not separable. The resolution of an infrared sensor is strongly dependent on the sensitivity of the sensor and vice versa. Therefore, a combined sensitivity and resolution performance parameter was developed called the minimum resolvable temperature difference (MRTD, or just MRT) that accurately describes the interrelationship for a particular sensor. MRT is used by many engineers and scientists, especially tactically oriented (fire control and targeting)

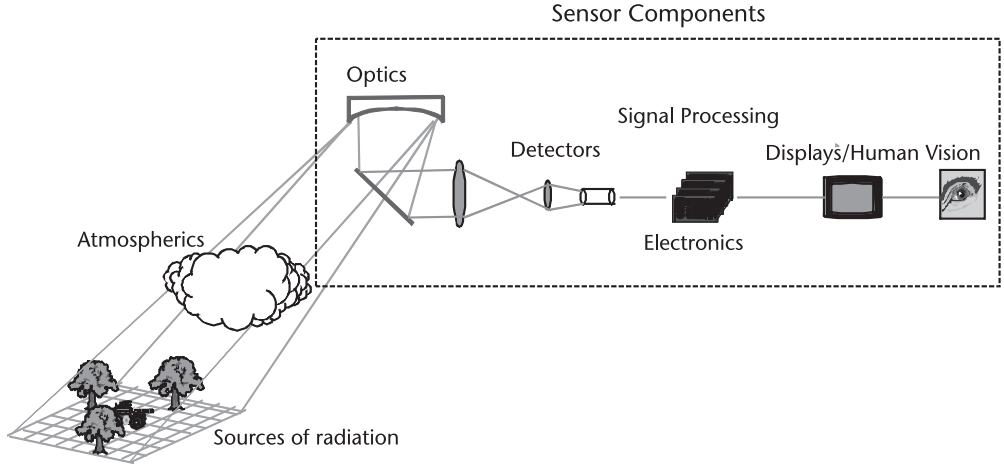


Figure 14.2 I^2R sensor system.

organizations, as their primary performance measure. Later, the system contrast threshold function (CTF) was developed to include a better estimate of system visual acuity that included the sensitivity and resolution of the system.

Many parameters are discussed in this chapter. The material is intended to pull together all of the sensor components presented in the previous chapters to culminate in a system-level sensor description. This is then used along with the target and atmospheric parameters to give a probability of object discrimination.

14.1 Sensitivity and Resolution

The two general parameters of performance for any I^2R or EO sensor are *sensitivity* and *resolution*. Sensitivity refers to how well a sensor can discriminate small radiance changes in the object space. Formally, sensitivity can be defined as the differential radiance that produces a unity SNR. For infrared sensors, this is usually described in terms of an equivalent blackbody temperature (i.e., the temperature of a ideal blackbody that delivers the prescribed radiance). Ideally, an infrared sensor would be able to resolve the radiant changes in targets and backgrounds to an infinite number of equivalent blackbody temperatures. However, system noise and sensor dynamic range eliminate any chance of infinite sensitivity. NETD is a good measure of sensitivity, but it is a gross value that lumps noise into an equivalent bandwidth. More important, sensitivity is a function of resolution. MRT describes sensitivity as a function of spatial frequency (i.e., resolution), producing a more comprehensive characterization.

Resolution refers to how well a sensor can see small spatial details in the object space. Resolution is characterized by the MTF of the system. However, the entire system MTF is not useful. Some MTF values decrease signals to less than the noise level of the system (for particular spatial frequencies). So, again, we find ourselves in a position where the true useful resolution of the system depends on noise (and therefore sensitivity). MRT and CTF were developed to address the need for a function relating the two.

Because sensitivity and resolution are related, changes in sensor resolution affect sensor sensitivity. Most of the possible changes in sensor components give conflicting changes in sensitivity and resolution. For example, an increase in a sensor's focal length (with all other sensor parameters held constant) may increase resolution and decrease sensitivity. Such a parameter change can give a sensitivity and resolution relationship similar to the one shown in Figure 14.3.

With such a parametric analysis, a sensor designer or analyst needs to determine an appropriate point to set the parameter. The decision may be a simple one because of the sensor use. For example, the sensor may be used to view rocket plumes where plenty of signal is present, but high resolution is desired. The process may be an iterative one requiring convergence to a compromise in requirements.

Not all parameters are conflicting in sensitivity and resolution. There are a few such as entrance pupil diameter that usually enhance both sensitivity and resolution. For a larger entrance pupil diameter, the resolution of a sensor may be increased because of a smaller diffraction spot (or psf) and the sensitivity of the sensor may be increased because of collection efficiency (i.e., flux throughput). Although

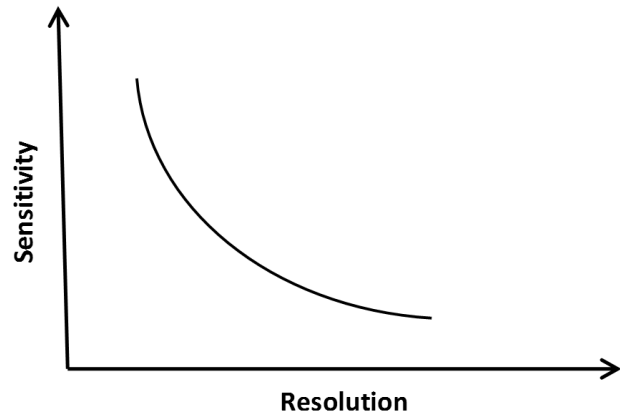


Figure 14.3 Sensitivity and resolution relationship.

this sounds like a great way to beat both parameters, imaging optics usually have a practical *f*-number lower limit of around 1, resulting in a limit on entrance pupil diameter. Also, optics with very low *f*-numbers tend to be accompanied by significant aberrations, so the resolution of the system may actually suffer. It is well known in the airborne sensor communities (both tactical and intelligence, surveillance, and reconnaissance [ISR]) that payload customers want sensors with infinite resolution, infinite sensitivity, and an infinite FOV. The bottom line is that sensor parameters involve significant trade-offs in sensitivity, resolution, and area coverage. The following sections provide an introduction to the sensitivity and resolution parameters of I^2R sensors, the basis of most trade-offs. We begin with the sensitivity parameters of NETD and noise equivalent irradiance (NEI).

14.2 Noise Equivalent Temperature Difference

NETD is a sensitivity parameter that is defined [3] as the target-to-background temperature difference in a standard test pattern that produces a peak-signal-to-rms noise ratio of unity at the output of a reference filter (recent measurements use the system noise bandwidth as the linearity filter) when the system views a test pattern. The NETD is a measure of the rms noise amplitude in terms of target-to-background temperature differentials. The NETD is typically measured with an oscilloscope on the output of a detector while the infrared imager views a square target of differential temperature ΔT that produces a signal-to-background differential voltage V_s. The noise voltage on the output of the detector is then measured with a true rms voltmeter V_n. The corresponding NETD of the detector signal is determined by

$$NETD = \frac{\Delta T}{V_s/V_n} \text{ [K]} \tag{14.1}$$

The overall frame NETD can then be taken as the average, or some other function [4], of the detector NETDs. A number of different techniques are actually used

to measure NETD that can lead to errors [4]. However, the basic concept is that NETD is a temperature equivalent signal that corresponds to system rms noise output.

The following derivation of NETD is provided by Lloyd [3] with the following assumptions:

1. Detector responsivity is uniform across the detector's sensitive area.
2. Detector detectivity is independent of other parameters in the NETD equation.
3. NETD equivalence is in blackbody signal levels.
4. The detector angular subtense and system *f*-number can be expressed by small angles.
5. Electronic processing introduces no noise.

Consider the sensor system in Figure 14.4, which views part of a large blackbody source. The blackbody source has a spectral radiance of

$$L(T,\lambda) = \frac{M(T,\lambda)}{\pi} \quad [\text{W/(cm}^2\text{-sr-}\mu\text{m)}] \tag{14.2}$$

The flux (or power) on the detector, whether the detector is scanned or staring can be expressed as

$$\Phi(T,\lambda) = \frac{M(T,\lambda)}{\pi}\frac{A_o}{R^2}(\alpha R)(\beta R)\tau_o(\lambda) = \frac{M(T,\lambda)}{\pi}A_o\alpha\beta\tau_o(\lambda) \quad [\text{W/}\mu\text{m}] \tag{14.3}$$

where $\alpha = a/f$ (horizontal detector angular subtense), $\beta = b/f$ (vertical detector angular subtense), A_o is the area of the optics entrance pupil, and $\tau_o(\lambda)$ is the transmission of the optics.

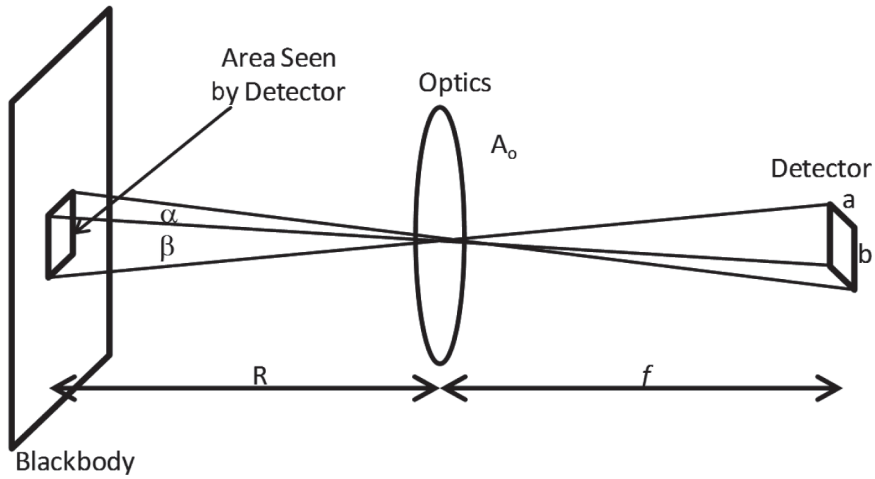


Figure 14.4 Blackbody flux to detector.

Because the detector signal responds to flux, we are interested in differential changes in flux with target temperature

$$\frac{\partial \Phi(\lambda)}{\partial T} = \frac{\alpha\beta}{\pi} A_o \tau_o(\lambda) \frac{\partial M(T,\lambda)}{\partial T} \quad [\text{W}/(\mu\text{m-K})] \tag{14.4}$$

The detector responsivity relates the flux $\Phi(\lambda)$ to the detector signal voltage by $V_n D^*(\lambda)/\sqrt{ab\Delta f}$, where $D^*(\lambda)$ is the detector detectivity and Δf is the detector electronics equivalent noise bandwidth. The change in voltage with the change in temperature can be written as

$$\frac{\partial V_s(\lambda)}{\partial T} = \frac{\alpha\beta}{\pi} A_o \tau_o(\lambda) \frac{V_n D^*(\lambda)}{\sqrt{ab\Delta f}} \frac{\partial M(T,\lambda)}{\partial T} \quad [\text{V}/(\mu\text{m-K})] \tag{14.5}$$

Integrating with wavelength yields an expression that is equal to $\Delta V_s/\Delta T$ for small signal approximations:

$$\frac{\Delta V_s}{\Delta T} = \frac{\alpha\beta A_o V_n}{\pi\sqrt{ab\Delta f}} \int_0^\infty \frac{\partial M(T,\lambda)}{\partial T} D^*(\lambda)\tau_o(\lambda)\, d\lambda \quad [\text{V/K}] \tag{14.6}$$

NETD occurs when $\Delta V_s/V_n$ is equal to 1. Rearranging (14.6) to solve for the ΔT when $\Delta V_s/V_n = 1$ gives

$$NETD = \frac{\pi\sqrt{ab\Delta f}}{\alpha\beta A_o} \frac{1}{\int_0^\infty \frac{\partial M(T,\lambda)}{\partial T} D^*(\lambda)\tau_o(\lambda) d\lambda} \quad [\text{K}] \tag{14.7}$$

This equation gives a generalized form of NETD that can be applied to a number of different systems. Equation (14.7) is an exact equation that is integrated in wavelength. NETD can be simplified further with a few assumptions or approximations:

1. The transmission of the optics $\tau_o(\lambda)$ is constant within the sensor band and zero outside the sensor band. An effective transmission for the band can be taken as

$$\tau_o(\lambda) = \frac{\int_0^\infty \frac{\partial M(T,\lambda)}{\partial T} D^*(\lambda)\tau_o(\lambda)\, d\lambda}{\int_0^\infty \frac{\partial M(T,\lambda)}{\partial T} D^*(\lambda)\, d\lambda} \tag{14.8}$$

2. Normalization of $D^*(\lambda)$ by the peak detectivity at wavelength Δ_p, $D^*(\lambda_p)$ and the definition of an *effective* change in object radiant emittance with temperature is

$$\frac{\Delta M}{\Delta T} \equiv \int_{\lambda_1}^{\lambda_2} \frac{\partial M(T,\lambda)}{\partial T} \frac{D^*(\lambda)}{D^*(\lambda_p)}\, d\lambda \tag{14.9}$$

Note that $D^*(\lambda_p)$ is a constant and must be multiplied by (14.9) in order to equal the integral in (14.7).

3. We can find $\frac{\partial M(T,\lambda)}{\partial T}$ from Planck's equation (Chapter 5) by

$$\frac{\partial M(T,\lambda)}{\partial T} = \frac{(2\pi hc^2)(hc/k)e^{hc/(\lambda kT)}}{\lambda^6 T^2 (e^{hc/(\lambda kT)}-1)^2} = \frac{M(T,\lambda)(hc/k)e^{hc/(\lambda kT)}}{\lambda T^2 (e^{hc/(\lambda kT)}-1)}$$
$$\approx M(T,\lambda)\frac{(hc/k)}{\lambda T^2} \tag{14.10}$$

for $e^{hc/(\lambda kT)} \gg 1$, which is reasonable for the imaging of terrestrial objects. The NETD equation is simplified further with the above approximations to

$$NETD = \frac{\pi\sqrt{ab\Delta f}}{\alpha\beta A_o \tau_o D^*(\lambda_p)\frac{\Delta M}{\Delta T}} = \frac{\pi\sqrt{ab\Delta f}}{\alpha\beta A_o \tau_o D^*(\lambda_p)\frac{(hc/k)}{T^2}\int_{\lambda_1}^{\lambda_2}\frac{M(T,\lambda)}{\lambda}\frac{D^*(\lambda)}{D^*(\lambda_p)}d\lambda} \tag{14.11}$$

Equation (14.11) is also a somewhat general expression for NETD. It can be simplified even further, however, if certain detector assumptions can be made. For a detector that operates at the theoretical limit of detectivity, the NETD can be extended for background-limited infrared photon (BLIP) detection systems. BLIP [5] occurs when the background photon flux collected by the detector is the dominant noise source in the system. Because the detectivity, or $D^*(\lambda)$, depends on the solid angle at which the detector collects the background radiation, a BLIP quantity called dee-double-star or dee-star-star $D^{**}_{BLIP}(\lambda)$is related to the detectivity

$$D^{**}_{BLIP}(\lambda) = \sqrt{\frac{\pi}{\Omega_{cs}}\eta_q}\, D^{**}_{BLIP}(\lambda) \tag{14.12}$$

where $D^{**}_{BLIP}(\lambda)$ term is normalized for unity quantum efficiency and an effective background collection angle of π. Consider the optical system and cold-shield solid angles shown in Figure 14.5.

The purpose of the cold shield is to block background radiation from striking the detector if the sources are not within the cold-shield solid angle. However, not

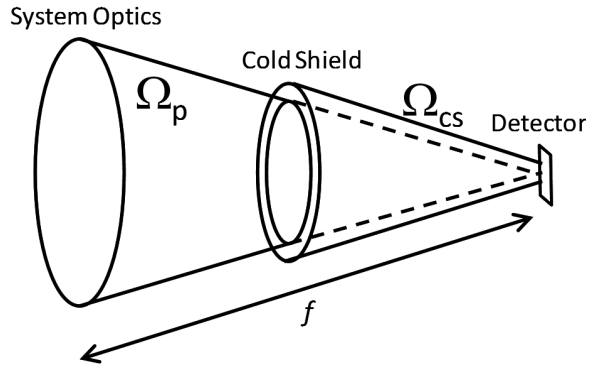


Figure 14.5 Cold-shield efficiency.

all cold shields are matched to the light input cone of the optics. The perfect cold-shield angle (for the small angle approximation) is

$$\Omega_p = \frac{\pi D^2}{4f^2} \tag{14.13}$$

where D is the input entrance pupil diameter and f is the system focal length. A cold-shield efficiency is taken to be

$$\eta_{cs} = \frac{\Omega_p}{\Omega_{cs}} \tag{14.14}$$

Substitution yields a detectivity of

$$D^*(\lambda) = 2\left(\frac{f}{D}\right)(\eta_{cs}\eta_q)^{1/2} D^{**}_{BLIP}(\lambda) \tag{14.15}$$

This gives an NETD of

$$NETD = \frac{2\sqrt{\Delta f}}{\sqrt{\alpha\beta\eta_{cs}\eta_q}\, D\, \tau_o D^{**}_{BLIP}(\lambda)\dfrac{\Delta M}{\Delta T}} \tag{14.16}$$

Finally, we have an NETD expression that can be used for gross sensitivity issues of an infrared sensor. Lloyd points out a number of deficiencies in using the NETD described here. Some of these are that (1) NETD is usually measured at the electronics output and not at the display, so it may not include all of the noise sources; (2) NETD is a measure of total in-band noise and the eye is not sensitive to all noise spatial frequencies in a uniform manner; and (3) NETD does not account for emissivities because blackbody radiation was assumed in the derivation. Recall that sensitivity is really a function of spatial frequency and NETD does not contain the spatial frequency dependence. It is well understood that MRT and system CTF are better measures of system performance. NETD is still used frequently, however, and as long as the limitations are understood, it can be a useful tool in system evaluation when used with other parameters. Typical values for system NETD range from around 0.01K to 0.07K.

For a perfect cold-shield design, (14.16) can be simplified with a few substitutions and some algebra. The substitutions include $a/f = \alpha$, $b/f = \beta$, $A_d = ab$, and $\Delta L = \Delta M/\pi$. Band-averaged detectivity is usually specified by detector manufacturers so that a useful form of NETD is

$$NETD = \frac{4}{\pi}\left[\frac{(f/\#)^2\sqrt{\Delta f}}{\tau_o\sqrt{A_d}D^*\dfrac{\Delta L}{\Delta T}}\right] \tag{14.17}$$

where $f/\#$ is the f-number equal to f/D and ΔL is the differential radiance. Note that the radiance difference is taken with respect to temperature, and D^* is the band-averaged detectivity in $\text{cm}\sqrt{\text{Hz}}/\text{W}$. Finally, $\dfrac{\Delta L}{\Delta T}$ is the band radiance derivative of Planck's law and can be approximated (e.g., for 300K optics paths, backgrounds)

Table 14.1 NETD Scanning System Parameters for Example 14.1

Entrance aperture diameter	20 cm
Focal length	35 cm
Horizontal FOV	4 deg
Vertical FOV	3 deg
Frame rate	30 Hz
Overscan ratio	1
Interlace	2:1
Number of detectors	180
Horizontal DAS	0.2 mrad
Vertical DAS	0.2 mrad
D^*	$5\times10^{10}\text{cm}\sqrt{\text{Hz}}/\text{W}$
Optics transmission	0.70
Scan efficiency	0.75

at 6.3×10^{-5} W/(cm^2-sr-K) for the 8- to 12-μm band and 6.7×10^{-6} W/(cm^2-sr-K) for the 3- to 5-μm band.

Example 14.1

Consider a lungwave scanning system that has the parameters given in Table 14.1 and determine the NETD.

The solution is given by determining the parameters of (14.17). First, the *f*/# is 1.75. The detector dimensions can be determined from the DASs and the focal length to give 70×70 μm. The detector area is 49×10^{-6} cm^2. If the bandwidth is matched to the system, the line scan time must be determined. With a 2:1 interlace, the detectors are scanned horizontally at a rate of 60 times per second. This means that a detector line is taken in 16.7 ms. With a scan efficiency of 0.75, the line time is 12.5 ms. There are 350 horizontal DASs across a line. Two samples or two integration times occur during the detector dwell so that effectively 700 independent signals are required from one line. Therefore, 700 signals per 12.5 ms are required for a sufficient bandwidth, or 55.9 kHz. The NETD is desired in the 8- to 12-μm waveband. Using (14.17), an NETD of 0.06K is obtained. Equation (14.17) does not take into account any improvement in NETD from the TDI process. For I^2R sensors with more than one row of scanning detectors that are summed in a TDI fashion, Shumaker [6] provides a NETC or a chain NETD that shows the improvement of the TDI chain:

$$NETC = \frac{NETD}{\sqrt{N_{TDI}}}$$

where NTDI is the number of TDI detectors in a chain. An example of this is the second-generation FLIRs with a four-detector TDI chain. The improvement here for the example above would be a factor of 2, or an NETC of 0.03K.

Example 14.2

Table 14.2 provides the midwave (3- to 5-μm) I^2R parameters for a staring system. Determine the NETD.

Again, the parameters in (14.17) must be determined. First, the *f*/# is calculated to be 3.1. The bandwidth can be approximated by $1/\tau_{int}$, where τ_{int} is the detector integration time. The integration time here is very similar to the detector dwell time

Table 14.2 Staring Sensor Parameters for Example 14.2

Entrance aperture diameter	17 cm
Focal length	53 cm
Horizontal FOV	2.0 deg
Vertical FOV	2.0 deg
Array size	$1{,}024 \times 1{,}024$
Frame rate	60 Hz
Fields per frame	1:1
Horizontal detector size	18
Vertical detector size	18
Horizontal pitch	18
Vertical pitch	18
D^*	$50 \times 10^{10} \text{cm}\sqrt{\text{Hz}}/\text{W}$
Optics transmission	0.8
Scan efficiency	1.0

in a scanning system. The bandwidth is then calculated to be 37.5 Hz. The area of a single detector is 18×18 μm, or 3.24×10^{-6} cm^2. The detectivity is given and the change in radiance with temperature is 6.7×10^{-6} W/(cm^2-sr-K) for the 3- to 5-μm band. Using these parameters gives the NETD at the output of a single detector. The NETD is calculated to be 0.015K.

Note that the example of the staring sensor gives a smaller but comparable NETD to that of the scanning sensor. This should not be taken as a direct comparison because the first system is a longwave system and the second is a midwave system. The change in radiance with temperature is much less in this band for terrestrial temperatures. The staring array is desirable because the integration time is larger, thus reducing the NETD values. If a system similar to the longwave example system were constructed in a staring array configuration, the NETD would be much smaller. Finally, the NETD value given for the staring array may not be a good indicator of system noise, because we have described the NETD for a single-detector output. Fixed-pattern noise causes the NETD to appear at larger values, so 3D noise may be a better system noise description.

14.3 EO Noise and Noise Equivalent Irradiance

The sensitivity of EO systems is limited by various noise contributors [7]. Some of the primary noise sources are as follows:

Shot noise: Caused by the discrete nature of photon to electron transitions, it occurs when photoelectrons are generated in the detection process and when dark current electrons are present.

Pattern noise: Noise that does not change significantly from frame to frame. This noise can be caused by detector construction nonuniformities.

Reset noise: Caused during the resetting of the sensing capacitor. This is the Johnson noise associated with the resistance in the resetting field effect transistor.

On-chip and off-chip amplifiers: Contributes both 1/*f* ("one over *f*") noise and white noise to the output signal. These noises are caused by the contacts and resistance in the amplifiers.

Quantization noise: The noise associated with the analog-to-digital converter's discrete output levels. These levels contribute an error (or noise) to the output signal.

Both the photon noise and the pattern noise are associated with the arrival of discrete light packets and the conversion of these packets to electrons. These noise contributors are small for low-light levels and increase with the amount of light received by the sensor. The other components listed are associated with the electronic readout of the CCD electron "wells." Over some integration time, a CCD detector will collect photoelectrons in a well that will be "read out" by the electronics. If the light level is high and the integration time is too long, the well can become filled, or saturated.

For low light levels (even for DTV systems), the noise is dominated by the electronic readout noise components. This includes reset noise, white noise, 1/*f* noise, and quantization noise. The readout noise is sometimes called the *noise floor* or *noise equivalent electrons*, referring to the number of equivalent electrons collected in a well over the integration time.

Holst [8] provides a simplification of system noise as the root-sum-square of the shot noise, readout noise, and the pattern noise:

$$< n_{sys} > = \sqrt{\left\langle n_{shot}^2 \right\rangle + \left\langle n_{readout}^2 \right\rangle + \left\langle n_{pattern}^2 \right\rangle} \quad \text{[electrons]} \tag{14.18}$$

where the brackets, <>, correspond to the ensemble average. For low light levels, the shot noise and the pattern noise are small and the readout noise is dominant. For high light levels, the shot noise and pattern noise may be larger than the readout noise.

Noise manifests itself as currents that reduce the overall performance of the system. The noise electrons described above contribute to the noise currents. The following descriptions are for currents corresponding to photon-electron (shot) noise, fixed-pattern noise, and readout noise.

Photon-Electron Shot Noise

The photon-electron (p-e) shot noise is a function of the amount of light collected from the target and the background. The average current can be taken as the average of the target (i_t) and background (i_b) currents:

$$i_{ave} = \frac{i_t + i_b}{2} \quad \text{[A]} \tag{14.19}$$

Also, the apparent contrast seen by the EO sensor can be approximated by

$$C \cong \frac{\left| i_b - i_t \right|}{i_b} \quad \text{[unitless]} \tag{14.20}$$

The p-e, or shot, noise is related to the average current such that the noise current is

$$i_{p-e}^2 = 2ei_{ave}\Delta f \quad [\mathrm{A}^2] \tag{14.21}$$

where e is the charge on an electron (1.6022×10^{-19} C) and Δf is the noise equivalent bandwidth. Equations (14.19) and (14.20) are used to simplify (14.21):

$$i_{p-e}^2 = (2 - \mathrm{C})ei_t\Delta f \quad [\mathrm{A}^2] \tag{14.22}$$

The current generated by the target flux is usually assumed to be of a larger value than that of the background current, but in practice, little difference between the detectability of objects is seen for positive or negative contrast. Therefore, we use i_b as the conversion factor because the background tends to encompass more the of sensor's FOV. Also, if a prestorage gain, G, is present and if horizontal and vertical readout efficiencies (e_h and e_v) are provided, then

$$i_{p-e}^2 = \frac{(2 - \mathrm{C})ei_b\Delta f G^2}{e_h e_v} \quad [\mathrm{A}^2] \tag{14.23}$$

Finally, the noise current density is related to the noise current by

$$J_{p-e}^2 = \frac{i_{p-e}^2}{2\Delta f} \quad [\mathrm{A}^2/\mathrm{Hz}] \tag{14.24}$$

$$= \frac{(2 - \mathrm{C})ei_b G^2}{2e_h e_v} \quad [\mathrm{A}^2/\mathrm{Hz}] \tag{14.25}$$

Note that the p-e noise is a function of target-to-background contrast and background current.

Fixed-Pattern Noise

The sensitivity of individual elements within a detector array can vary significantly. Two common causes are *fabrication artifacts* and *material impurities*. Fixed-pattern noise is not temporal in nature (it can, however, change with scene brightness) and is fixed when viewing a static scene. However, it is converted to temporal noise in the readout process. Fixed-pattern noise is also a function of background current, with the current density identified as

$$J_{fp}^2 = \frac{[Mi_b]^2 t_f}{N_x N_y} \quad [\mathrm{A}^2/\mathrm{Hz}] \tag{14.26}$$

where M is the standard deviation of the fixed-pattern noise. It is sometimes called the *fixed-pattern noise modulation*. The N_x and N_y terms represent the number of pixels in the x and y directions, respectively, and tf is the frame integration time (inverse of the frame rate). The current density can be used to determine the fixed-pattern noise current given an "information bandwidth":

$$\Delta f_I = \frac{N_x N_y}{2t_f} \quad [\mathrm{Hz}] \tag{14.27}$$

Using the noise current and current density relationship of (14.26), the noise current density is

$$i_{fp}^2 = (Mi_b)^2 \quad [\text{A}^2] \tag{14.28}$$

Note that the fixed-pattern noise current is a function of the fixed-pattern noise modulation and the background current.

Readout (Floor) Noise

The readout noise, or floor noise, is the catch-all term for all other noise sources that are left over. It includes on-chip amplifier, Johnson, and off-chip amplifier noise. It is usually specified by some equivalent $n_{readout}$ noise electrons collected by a detector over some integration time. The mean-squared floor noise current is given by

$$i_{readout}^2 = \frac{4\Delta f_l^2 (n_{readout} e)^2}{(e_h e_v)^2} \quad [\text{A}^2] \tag{14.29}$$

and the noise current density is given by

$$J_{readout}^2 = \frac{2\Delta f_l (n_{readout} e)^2}{(e_h e_v)^2} \quad [\text{A}^2/\text{Hz}] \tag{14.30}$$

The readout noise is not usually a problem for a sensor when the scenario contains a high background current, but it can easily be the limiting factor for sensors in low light levels.

Total Noise Current

The combined noise in an EO system can be approximated by the following simplified noise model:

$$i_{noise} = \sqrt{i_{p-e}^2 + i_{fp}^2 + i_{readout}^2} \quad [\text{A}] \tag{14.31}$$

Recall that the p-e and fixed-pattern noises were functions of the background current such that a typical system noise is given as a function of the background light level as shown in Figure 14.6.

The system noise shown is limited by readout noise at low light levels and by p-e and fixed-pattern noise at higher levels. Also, even though the noise level increases at the higher light levels, the target-to-background signal is also increasing. Usually, the SNR increases until p-e and fixed-pattern noise become dominant. Then, the SNR approaches a constant value for higher light levels.

Example 14.3

Determine the total EO sensor noise current for a 480 × 640 CCD array with the following characteristics: a 6.4-μA background current, a target-to-background contrast of 0.33 (33%), a frame rate of 30 Hz, horizontal and vertical readout efficiencies of 0.9, and a fixed-pattern modulation of 0.0005. Assume that the readout noise is 120 electrons and the prestorage gain is 1. Also, determine the limiting noise contributor (from readout noise, fixed-pattern noise, and shot noise). The so-

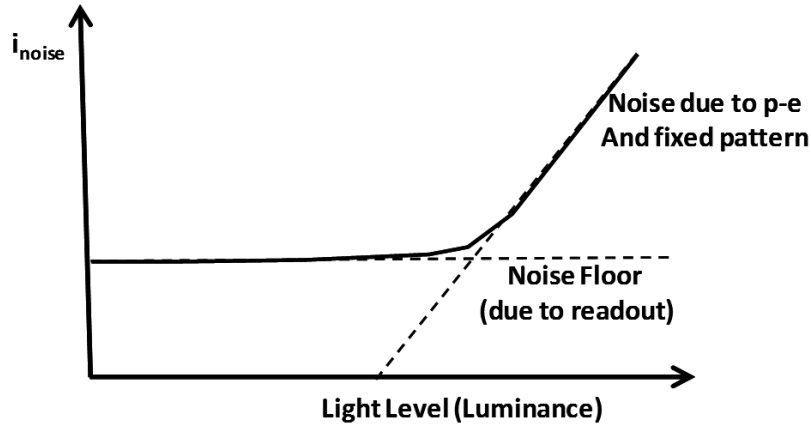


Figure 14.6 Total EO system noise current.

lution is provided with the calculation of each noise contributor. First, Δf is assumed to be 30 Hz and Δf_l = (640)(480)/(2 × 1/30) or 4.61 MHz. The solution is:

$$i_{p-e}^2 = \frac{(2-0.33)(1.6\times10^{-19}C)(6.4\times10^{-6})(1/30)}{(0.9)(0.9)} = 7.04\times10^{-26} \quad [\text{A}^2]$$

$$i_{fp}^2 = [(0.0005)(6.4\times10^{-6})]^2 = 1.02\times10^{-17} \quad [\text{A}^2]$$

$$i_{readout}^2 = \frac{2(4.6\times10^{6})[(120)(1.6\times10^{-19})]^2}{[(0.9)(0.9)]^2} = 2.38\times10^{-20} \quad [\text{A}^2]$$

From the values calculated above, the fixed-pattern noise provides the largest noise current. The total rms noise current can be determined by adding the above contributors and taking the root in a root-sum-squared fashion to give an i_{noise} of 3.19 rms nanoamperes.

Noise Equivalent Irradiance

The NEI of an EO system is similar to the NETD parameter for the I²R system. In a manner similar to NETD, we consider the exitance (recall that exitance includes reflected flux) of a surface, $M_q(\lambda)$, in photon/s-cm²-μm, as shown in Figure 14.7. Assuming a Lambertian surface, the radiance is

$$L_q(\lambda) = \frac{M_q(\lambda)}{\pi} \quad [\text{photon/s-cm}^2\text{-sr-}\mu\text{m}] \tag{14.32}$$

The amount of area seen by the detector is

$$A_{src} = \frac{a}{f}R\frac{b}{f}R = \frac{ab}{f^2}R^2 \quad [\text{cm}^2] \tag{14.33}$$

The intensity of the source area seen by the detector is

$$I_q(\lambda) = \frac{M_q(\lambda)}{\pi}\frac{ab}{f^2}R^2 \quad [\text{photon/s-sr-}\mu\text{m}] \tag{14.34}$$

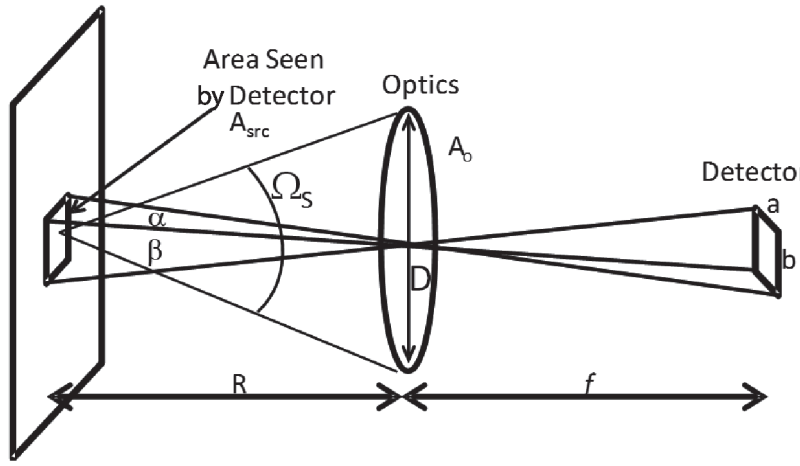


Figure 14.7 NEI nomenclature.

The total amount of the intensity that enters the aperture of the sensor corresponds to the amount of flux that falls on the detector:

$$\Phi_q = \int_{\lambda_1}^{\lambda_2} \Omega_s \tau_{optics}(\lambda) I_q(\lambda)\, d\lambda = \frac{D^2 ab}{4f^2} \int_{\lambda_1}^{\lambda_2} \tau_{optics}(\lambda)\, M_q(\lambda)\, d\lambda \quad \text{[photon/s]} \quad (14.35)$$

where $\Omega_s = \dfrac{\pi D^2}{4f^2}$ is the solid angle of the sensor. Over an integration time t_{int} and a quantum efficiency η_q, the number of electrons collected is

$$N_q = \frac{D^2 ab \eta_q t_{int}}{4f^2} \int_{\lambda_1}^{\lambda_2} \tau_{optics}(\lambda) M_q(\lambda)\, d\lambda \quad \text{[photons]} \quad (14.36)$$

We know the number of noise electrons, n_{sys}, so with the average optical transmission, τ_{optics}, we can rearrange (14.36) to determine the NEI:

$$NEI = \frac{4f^2 n_{sys}}{D^2 \eta_q ab \tau_{optics}} \quad \text{[photon/s-cm}^2\text{]} \quad (14.37a)$$

where the area of the detector, A_d, is the product of a and b. NEI corresponds to the exitance of the surface that gives a SNR of unity.

There is a similar quantity to *NEI* that gives an EO sensor's sensitivity in terms of the photometric irradiance (illuminance) at the entrance aperture of the sensor. Note that this is different than an exitance like NEI. The assumptions here are that the illuminance at the entrance aperture was caused by a distant point source. The derivation is very similar to the one described above (it begins at the entrance aperture of the sensor and is assumed to be a plane wave).

The minimal photometric signal in terms of illuminance [8] is

$$E_v = \frac{n_{readout} hc}{\eta_q \Delta\lambda \tau_{optics} t_{int}} \frac{683}{A_o} \int_{0.38}^{0.75} \frac{V(\lambda)}{\lambda} d\lambda \quad \text{[lx]} \quad (14.37b)$$

where $V(\lambda)$ is the photopic efficiency, h is Planck's constant, c is the speed of light, η_q is the detector quantum efficiency, $\Delta\lambda$ is the spectral band, and τ_{optics} is the band-averaged optical transmission. Finally, t_{int} is the integration time, $n_{readout}$ is the readout noise in electrons, and A_o is the entrance aperture area.

The sensitivity of most commercial camcorders is specified with the minimal photometric signal. A typical home-use camcorder will have a sensitivity of less than a lux. The problem with the photometric specification is that it is applied to sensors that are not closely matched to the human eye. For a sensor with a different spectral response, a photometric specification can be meaningless.

14.4 Three-Dimensional Noise

Three-dimensional noise [9–11] has become a more common method of characterizing infrared system noise. NETD is limited in that it only characterizes temporal detector noise, where three-dimensional noise characterizes both spatial and temporal noises that are attributed to a wide variety of sources. In fact, 3D noise was successfully integrated into the U.S. Army NVESD's FLIR92 MRTD sensor model and parts of it were integrated into later models. Consider the successive frames of acquired noise given in Figure 14.8.

A directional average is taken within the coordinate system shown in order to obtain eight parameters that describe the noise at the system's output. The noise is then calculated as the standard deviation of the noise values in the directions that were not averaged. The parameters are given in Table 14.3, where the subscript that is "missing" gives the direction that was averaged. The directional averages are converted to equivalent temperatures in a manner similar to NETD. The result is a set of eight noise parameters that can be used as analytical tools in sensor design, analyses, testing, and evaluation. The majority of these parameters cannot be

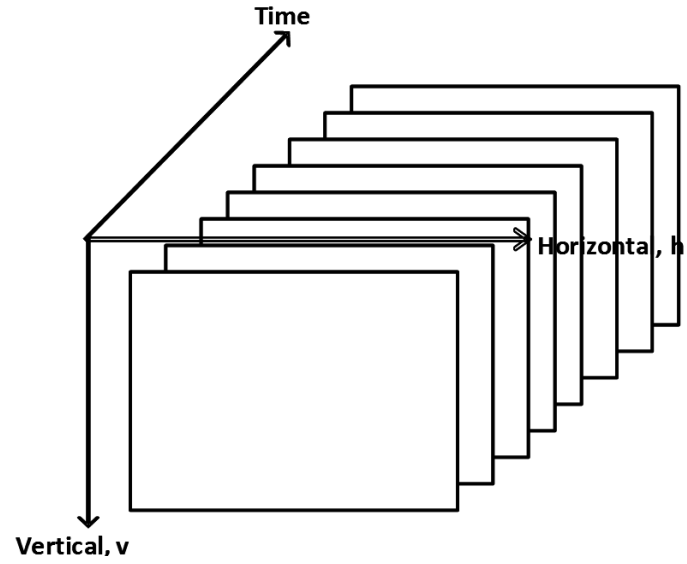


Figure 14.8 Three-dimensional noise coordinates.

Table 14.3 Three-Dimensional Noise Components

Noise	*Description*	*Source*
σ_{tvh}	Random spatiotemporal noise	Detector temporal noise
σ_{tv}	Temporal row noise, line bounce	Line processing, 1/*f* readout
σ_{th}	Temporal column noise, column bounce	Scan effects
σ_{vh}	Random spatial noise	Array detector nonuniformity
σ_{v}	Fixed-row noise	Line detector nonuniformity
σ_{h}	Fixed-column noise	Scan effects
σ_{t}	Frame-to-frame noise	Frame processing
S	Mean of all noise components	

From [9].

calculated like NETD with the exception of σ_{tvh}, which is similar to NETD; in fact, σ_{tvh} is actually identical to NETD with the exception that the actual system noise bandwidth is used instead of the reference filter bandwidth. The other noise parameters can only be measured to determine the infrared sensor artifacts. In the FLIR92 infrared sensor model, reasonable estimates are made for these parameters based on a large database of historical measurements. The measurements were conducted on both scanning and staring system noise parameters.

If all of the noise components are considered statistically independent, an overall noise parameter can be given at the system output as

$$\Omega = \sqrt{\sigma_{tvh}^2 + \sigma_{th}^2 + \sigma_{tv}^2 + \sigma_{vh}^2 + \sigma_{h}^2 + \sigma_{v}^2 + \sigma_{t}^2} \tag{14.38}$$

The frame-to-frame noise is typically negligible, so it is not included in most noise estimates.

The 3D noise can be expanded further to include the perceived noise with eye and brain effects in the horizontal and vertical directions. Composite system noise (perceived) in the horizontal direction can be given by

$$\Omega_h = \left[\sigma_{tvh}^2 E_t E_v(\xi) E_h(\xi) + \sigma_{vh}^2 E_v(\xi) E_h(\xi) + \sigma_{th}^2 E_t E_h(\xi) + \sigma_{h}^2 E_h^2(\xi)\right]^{1/2} \tag{14.39}$$

where E_t, $E_v(\xi)$, and $E_h(\xi)$ are the eye and brain temporal integration, vertical spatial integration, and horizontal spatial integration, respectively. In the vertical direction, the composite noise is given by

$$\Omega_v = \left[\sigma_{tvh}^2 E_t E_v(\eta) E_h(\eta) + \sigma_{vh}^2 E_v(\eta) E_h(\eta) + \sigma_{tv}^2 E_t E_v(\eta) + \sigma_{v}^2 E_v^2(\eta)\right]^{1/2} \tag{14.40}$$

Note that the noise terms included in each perceived composite signal correspond to only those terms that contribute in that particular direction.

Scanning systems show a wide variety of noise values. Three different estimates of 3D noise values corresponding to low-, moderate, and high-noise systems are provided in Table 14.4. Staring arrays have been dominated by random spatial noise, so a single noise model is used. Estimates provided are based on a measurement

Table 14.4 Three-Dimensional Estimates Based on Historical Measurements

Noise Term	*Scanning Low Noise*	*Scanning Moderate Noise*	*Scanning High Noise*	*Staring Noise*
σ_{vh}	0	0	0	$0.40\sigma_{tvh}$
σ_{tv}	$0.25\sigma_{tvh}$	$0.75\sigma_{tvh}$	$1.0\sigma_{tvh}$	0
σ_{v}	$0.25\sigma_{tvh}$	$0.75\sigma_{tvh}$	$1.0\sigma_{tvh}$	
σ_{th}	0	0	0	0
σ_{h}	0	0	0	0

database at the U.S. Army's NVESD for infrared system characterizations. These estimates are given in Table 14.4 in terms of a percentage of the random spatiotemporal noise.

14.5 Modulation Transfer Function

The MTF of a system is a primary measure of the overall system resolution. The system MTF gives the transfer of input spatial frequencies in a manner similar to the transfer function in electronic systems for temporal frequencies. Usually, the MTF can be considered separable in the horizontal and vertical directions. For example, the horizontal MTF of a system includes all of the system horizontal components:

$$MTF_{sys}(\xi) = H_{optics}(\xi)H_{detector}(\xi)H_{electronics}(\xi)H_{display}(\xi)H_{eye}(\xi) \tag{14.41}$$

where the eye transfer function used to be included in the MTF for use in the MRT models (like FLIR92), but it is not included in the newer system CTF models (like NVTherm or NVIP). Also note that the atmospheric MTF of the turbulence is not included in the system MTF, but it can be cascaded with the system MTF if the overall transfer of an imaging system within a particular scenario is desired. The addition of atmospheric MTF would allow the output image spectrum (and resulting image) to be determined from an input object spectrum (or object scene). The atmospheric MTF is usually left out, however, because the sensor characteristics may need to be scene independent or a relative performance between two sensors is needed. Any atmospheric MTF that is selected would limit the sensor characterization to that scenario. The horizontal sensor MTF given is only for the part of the system within the dashed box in Figure 14.2. In addition, note that the atmospheric MTF is a function of range, so the MTF must be changed as the sensor is positioned closer to the target.

The form of the vertical MTF is identical to (14.41), with η substituted for ξ, and the vertical MTF components may be different from those in the horizontal directions. Therefore, two MTFs are calculated for the sensor. MTFs that are circularly symmetric, such as the diffraction component of the optics MTF, may not be separable in Cartesian coordinates. A slice through the circular function in the horizontal direction, however, can give a reasonable estimate for how the component MTF behaves in that direction. This one-dimensional slice function can be used as horizontal and vertical MTF estimates.

At each spatial frequency, the MTF value for each component in the system is multiplied to determine system MTF at that particular spatial frequency. Figure 14.9 shows an example, where the component MTFs are shown. They are point-by-point multiplied to determine the overall system MTF. Also, one should not assume that the system cutoff frequency is where the MTF approaches zero. The system noise level provides a noise floor that limits the usefulness of small MTF values. For example, a high spatial frequency signal may propagate through the system MTF with a transfer of 10% (compared with dc light levels), but the noise may drown out signals that have been reduced to this output level.

To illustrate the actions of the system MTF on the resolution of an image, consider Figure 14.10. An input scene is shown with variable frequency, where the frequency increases linearly from left to right. The amplitude of the signal in the horizontal direction is constant for all the frequencies shown so the spectrum is uniform over the frequency range. The spectrum is multiplied by the sensor MTF to determine the sensor's output image spectrum. The corresponding output spectrum is shown. Note that while the amplitude for all frequencies on the input is uniform, the high-frequency amplitudes suffer on the output image of the sensor. Reduced signal amplitudes on the output correspond to smaller modulation at the output of the sensor. Recall that the same effect can be seen on the output image by convolving the input scene with the sensor psf.

Sometimes $MTF_{sys}(\xi)$ is broken into two parts for undersampled imaging systems: $MTF_{pre}(\xi)$ for a presample MTF and $MTF_{post}(\xi)$ for a postsample MTF. It is important to use these two MTFs separately to determine aliasing content as described in the linear systems chapter. However, for well-sampled imagers, it is usually customary to assume that the entire system is LSI and cascade all of the MTF components into an overall system MTF.

Finally, it is worth considering two additional MTF components that are frequently very important and are not covered in the previous component chapters: vibration and motion (or smear). Vollmerhausen et al. [12] provide a detailed

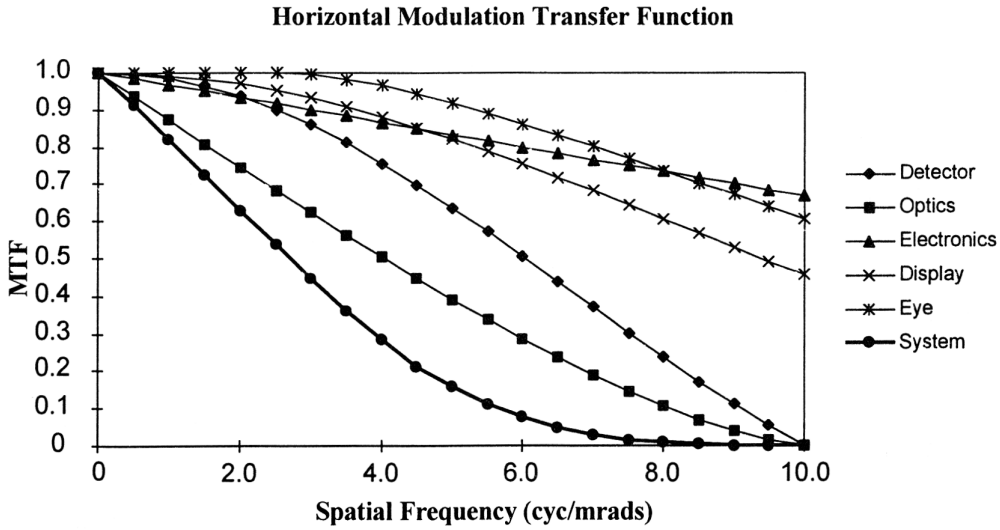


Figure 14.9 Modulation transfer function.

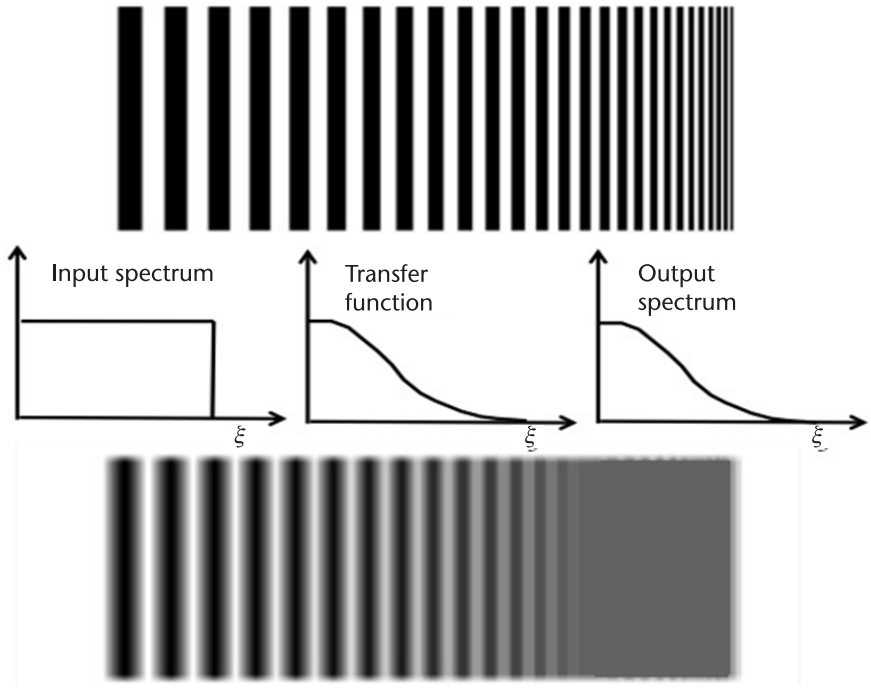


Figure 14.10 MTF response.

description of how these two MTFs impact system performance. Many times, the vibration blur can be described by a Gaussian MTF of

$$H_{vib}(\xi) = exp(-\beta^2\xi^2) \tag{14.42}$$

where β is the root-mean-squared blur radius of the vibration assuming that the Gaussian is a reasonable description of the vibration spectra within the sensor frame integration time. As always, ξ is the spatial frequency in cycles per milliradian. For a more general treatment of vibration and complex spectral, see Ref. [12].

Linear motion that causes smear can also be modeled as an MTF component that is part of the presample blur. The MTF in this case is

$$H_{motion}(\xi) = sinc(\phi\tau\xi) \tag{14.43}$$

where ϕ is the angular motion rate in milliradians per second and τ is the detector integration time of the imager. Both vibration and motion can significantly impact or limit the performance of the imager on certain platforms.

14.6 Minimum Resolvable Temperature Difference (Including 2D MRT)

The minimum resolvable temperature difference [13–15] is defined [15] as the differential temperature of a four-bar target that makes the target just resolvable by a

particular sensor. MRT is a sensor parameter that is a function and not just a value. It provides sensor sensitivity as a function of four-bar target frequency (i.e., resolution). The idea here is that the detection, recognition, and identification criteria of a target can be given in terms of a four-bar target spatial frequency. The response of the sensor is also given as a function of a four-bar frequency response. The sensor performance is then determined using the combination of sensor response and the target characteristics.

The MRT is measured using 7:1 aspect ratio targets as shown in Figure 14.11. The target-to-background differential temperature is varied until the four bars are just resolvable to a trained observer. Sometimes the differential temperature is increased from a small value until the target is just resolved and then decreased from a large value until the target is just resolved. The average of the two differential temperatures is then taken as the MRT for the particular target frequency. The same procedure is performed using negative contrast targets (the temperature of the target is cooler than the temperature of the background) to nullify test offsets. The average of the positive and negative magnitudes is then taken as the MRT. The differential temperature is plotted for each target frequency, where the collection of data points is the MRT. Note that the MRT curve appears to be related to the MTF curve in an inverted manner. This is indeed the case, as the subsequent discussion shows.

Lloyd [3] describes the four most important elements of image quality as sharpness, graininess, contrast rendition, and interference by artifacts. MRT became a widely accepted method for describing these characteristics in I^2R sensor performance because they are dominated by spatial resolution and thermal sensitivity. The derivation of MRT requires a number of assumptions. Those included in the NETD derivation are still considered applicable along with the following:

1. The eye integration time can be approximated between 0.1 and 0.2 sec.
2. The effect of spatial filtering in the eye for the display of a periodic square bar target of frequency f_T can be approximated by a matched filter for a single bar at $H(\xi) = sinc\left(\frac{\xi}{2f_t}\right)$.

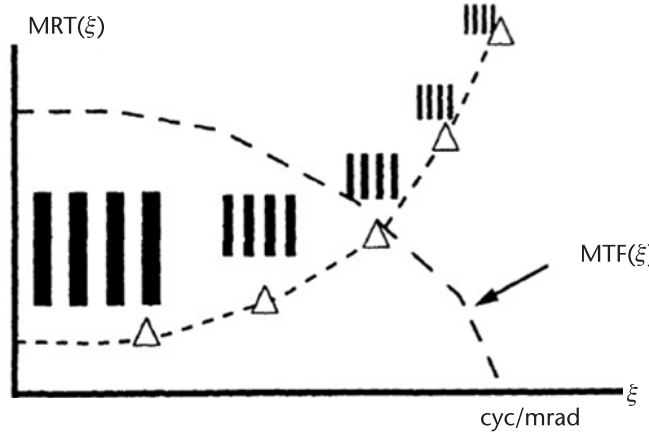


Figure 14.11 MRT curve.

3. The electronic processing and monitor are assumed noiseless.
4. The system can be considered an LSI system, where the spatial frequency transfer can be described by the system MTF.

The detailed derivation of MRT is found in Lloyd [3], where it is developed using a perceived SNR for human vision of a four-bar target. The model assumes the matched filter listed in the above assumptions and also that the first harmonic of the four-bar target frequency is passed by the system (i.e., the square wave response is converted to a sine wave response). The perceived SNR is also found to improve with the number of time samples per eye integration time.

The MRT equation has undergone numerous changes and has been approximated in many forms. A particular version of the MRT model used frequently was the FLIR92 model by the U.S. Army's NVESD. It is not the current NVESD model, but it is useful to cover from a theoretical past point of view and it is sometimes used with older systems even today. The horizontal MRT of an I^2R sensor is given by

$$MRT_z(\xi) = \frac{\pi^2}{8}\frac{SNRT\sigma_{tvh}k_z(\xi)}{MTF(\xi)}\left[E_t E_{h_z}(\xi)E_{v_z}(\xi)\right]^{1/2} \quad \text{[K]} \tag{14.44}$$

where z is the direction of the MRT (h for horizontal and v for vertical). $SNRT$ is the eye signal-to-noise threshold required to resolve an MRT target and is usually set to around 2.5. The random spatiotemporal noise σ_{tvh} is similar to the NETD with the exception of the bandwidth reference filter (the actual noise bandwidth of the electronics is used). Therefore, the random spatiotemporal noise is given by

$$\sigma_{tvh} = \frac{4f^2\Delta f}{\pi\tau_{optics}D^2\sqrt{ab}\left[\int_0^\infty D^*(300,\lambda)\,\frac{\partial M(\lambda)}{\partial T}d\lambda\right]} \quad \text{[K]} \tag{14.45}$$

Note that the random spatiotemporal noise is related to NETD only by the reference filter bandwidth used to measure NETD and the infrared sensor electronics bandwidth:

$$\sigma_{tvh} = NETD\frac{\sqrt{\Delta f}}{\Delta f_r} \tag{14.46}$$

If a reference filter is not used on the output of the system during the NETD measurement, which is a common practice in many laboratories, the two noise parameters are identical.

The eye–brain temporal and spatial integration effects are given by E_t, $E_h(\xi)$ and $E_v(\xi)$. The temporal eye and brain integral is approximated by

$$E_t = \frac{\alpha_t}{F_R t_{eye}} \tag{14.47}$$

where F_R is the frame rate and t_{eye} is the integration time of the eye. Frame rates are typically 30 Hz and eye integration times are around 0.1 sec for an observer viewing a display brightness of 0.15 mL. Smaller eye integration times are appropriate for higher display brightness. The temporal sample correlation factor α can usually be set to 1.

The eye–brain spatial integrations describe the human visual filtering of the noisy bar targets. These integrals are a little different for a scanning system and a staring system. For the scanning system horizontal MRT, the horizontal and vertical integrations are given by

$$E_{h_h}(\xi) = \frac{\nu}{\Delta f}\int_0^\infty S(\alpha)H_{NF_h}^2(\alpha)sinc^2\left(\frac{\alpha}{2\xi}\right)d\alpha \tag{14.48}$$

and

$$E_{v_h}(\eta) = \frac{DAS_v}{N_{os}}\int_0^\infty H_{NF_v}^2(\alpha)sinc^2\left(\frac{7\alpha}{2\eta}\right)d\alpha \tag{14.49}$$

The scan velocity ν in (14.48) can be estimated by first calculating the detector dwell time (the average time it takes a detector to scan across a point in the sensor's FOV)

$$t_D = \frac{N_v DAS_h DAS_v \eta_{sc}}{FOV_h FOV_v F_R N_{os}} \quad \text{[sec]} \tag{14.50}$$

where N_v is the number of detectors in the vertical direction; the detector angular subtenses and the sensor FOVs are given in milliradians; the scan efficiency, η_{sc}, is unitless; and N_{os} is the overscan ratio (or the number of samples per detector angular subtense in the vertical direction). The scan velocity is simply

$$\nu = \frac{DAS_h}{t_D} \quad \text{[mrad/s]} \tag{14.51}$$

The system noise bandwidth for scanning systems is considered the bandwidth at the output port of the electronics (does not include the display monitor):

$$\Delta f = \int_0^\infty S(\alpha)H_{TPF}^2(\alpha)\,sinc^2(\alpha t_i)\,d\alpha \quad \text{[Hz]} \tag{14.52}$$

where $S(\alpha)$ is the noise power spectrum, $H_{TPF}(\alpha)$ is the temporal postfilter transfer function, and ti is the integration time of the detector. The dummy variable α in (14.52) has units in hertz. The noise power spectrum can be estimated as a combination of white noise and 1/*f* noise:

$$S(f_e) = 1 + \frac{f_k}{f_e} \tag{14.53}$$

where f_k is the 1/*f* noise knee frequency. The temporal postfilter transfer function, $H_{TPF}(f_e)$, is a combination of all transfer functions between the detector output and the electronics output. This transfer function includes the finite response of the detector, the electronics lowpass filter, the electronics highpass filter, and any electronics boost:

$$H_{TPF}(f_e) = H_{dt}(f_e)H_{elp}(f_e)H_{ehp}(f_e)H_{eb}(f_e) \tag{14.54}$$

The effect of the electronics highpass filter is typically minimal. The $H_{dt}(f_e)$ and $H_{elp}(f_e)$ both represent lowpass filters, but with the detector 3-dB cutoff and the electronics 3-dB cutoff, respectively. The electronics boost transfer function is

$H_{eb}(f_e)$. Note that $H_{TPF}(f_e)$ only applies in the horizontal, or scan, direction. Both (14.53) and (14.54) can be written as a function of temporal frequency (as shown) or spatial frequency with the conversion, $f_e = v\xi$. The conversion is necessary in (14.48) and (14.49). Usually, the transfer functions of the filters in $H_{TPF}(f_e)$ are not as limiting as the *sinc* in (14.52) because of the detector integration time. However, the *sinc* term should only be present in (14.52) for detectors that have integration circuits. For example, all second-generation FLIRs would have the *sinc* term; however, most first-generation FLIRs that include the EO mux would not include this term. For detectors that do not integrate, the system noise bandwidth is determined primarily from the electronics lowpass filter.

Now that we have developed the lower level parameters, we can evaluate the eye–brain spatial integrals for the scanning system horizontal MRT. These spatial integrals are the response of the eye–brain to a noisy bar target at the display. The integral is one of a matched filter where the eye integrates the noise over a bar that is matched to that of the four-bar target. The bar height is 7 times larger in the vertical direction than in the horizontal direction, thus, the 7 is in the frequency response of the vertical *sinc* term.

The noise filter H_{NF} is the detector-to-eye filter that includes all transfer functions from the detector to the observer's eye. In the horizontal direction,

$$H_{NF_h}(\xi) = H_{TPF}(\xi)H_{SPF_h}(\xi)\,sinc(\xi v t_i) \tag{14.55}$$

where $H_{SPF}(\xi)$ is the spatial postfilter that includes any spatial processing between the detector output and the human eye. It usually includes the monitor MTF (typically a CRT with accounting for the spot-size and sample-and-hold), the human eye MTF, and any digital electronics spatial processing. In the vertical direction, the noise filter is simply

$$H_{NFv}(\xi) = H_{SPFv}(\xi) \tag{14.56}$$

where the spatial postfilter in the vertical direction includes the monitor (sample-and-hold does not apply to the vertical direction), the human eye, and any digital electronics processing.

As shown, many parameters affect the spatial integrals described by (14.48) and (14.49). The scanning system vertical MRT spatial integrals are similar:

$$E_{h_h}(\xi) = \frac{v}{\Delta f}\int_0^\infty S(\alpha)H_{NF_h}^2(\alpha)sinc^2\left(\frac{7\alpha}{2\xi}\right)d\alpha \tag{14.57}$$

and

$$E_{v_h}(\eta) = \frac{DASv}{N_{os}}\int_0^\infty H_{NF_v}^2(\alpha)\,sinc^2\left(\frac{\alpha}{2\eta}\right)d\alpha \tag{14.58}$$

The only difference here is that the MRT bars are rotated to the vertical direction. That is, the bars are longer in the horizontal direction by a factor of 7.

The last MRT parameter is $k_z(\xi)$, the noise correction factor. The horizontal noise correction factor can be calculated using the 3D noise parameters given in the previous section. This parameter provides the effects of the other-than-random spatial noise. For a rough estimate of a sensor design MRT, this parameter can be

set to 1. However, the previous section gave estimates for the various types of sensors based on real data. These estimates should be considered for a detailed design of a sensor. The horizontal correction factor is

$$k_h(\xi) = \left[1 + \frac{1}{E_t}\left(\frac{\sigma_{vh}}{\sigma_{tvh}}\right)^2 + \frac{1}{E_{v_h}(\xi)}\left(\frac{\sigma_{th}}{\sigma_{tvh}}\right)^2 + \frac{1}{E_t E_{v_h}(\xi)}\left(\frac{\sigma_h}{\sigma_{tvh}}\right)^2\right]^{1/2} \tag{14.59}$$

and the vertical correction factor is

$$k_v(\xi) = \left[1 + \frac{1}{E_t}\left(\frac{\sigma_{vh}}{\sigma_{tvh}}\right)^2 + \frac{1}{E_{h_v}(\eta)}\left(\frac{\sigma_{tv}}{\sigma_{tvh}}\right)^2 + \frac{1}{E_t E_{h_v}(\eta)}\left(\frac{\sigma_v}{\sigma_{tvh}}\right)^2\right]^{1/2} \tag{14.60}$$

The estimates for these values can be found for scanning and staring systems in the previous section.

Example 14.4

A longwave second-generation, parallel scan I^2R sensor has the parameters stated in Table 14.5. Determine the horizontal MRT.

Table 14.5 Second-Generation MRT Parameters for Example 14.4

Spectral band	7.5–10.0
Focal length	25 cm
Aperture diameter	12.5 cm
Average optical transmission	0.7
Horizontal/vertical aspect	1.33
Frame rate	30 Hz
Fields per frame	1
Horizontal detector dimension	40
Vertical detector dimension	40
Peak D^*	10E10 Jones
D^* wavelength	[7.50 8.00 8.5 9.0 9.50 10.0]
Normalized D^*	[0.75 0.80 0.85 0.9 0.95 1.0]
Detector integration time	10 μs
1/f knee frequency	100 Hz
Detectors in TDI	4
Number of vertical detectors	480
Samples per DASv	2
Samples per DASh	2
Detector 3-dB response frequency	500,000 Hz
Scan efficiency	0.75
Electronics highpass 3-dB frequency	1 Hz
Electronics highpass filter order	1
Electronics lowpass 3-dB frequency	100,000 Hz
Electronics lowpass filter order	1
No boost filter	
Number of display lines	480
Display brightness	10 ftL
Display height	15.24 cm
Display viewing distance	30.0
Threshold SNR	2.5
Eye integration time	0.1 sec

The MRT for the second-generation FLIR is given in (14.44). The first step in evaluating this equation is to determine the system MTF. The system MTF comprises the prefilter rollup transfer functions, the temporal postfilter transfer functions, and the spatial postfilter transfer functions. The prefilter rollup transfer functions are shown in Figure 14.12. The optics transfer function is the incoherent transfer function corresponding to diffraction. We are assuming (since the specifications did not state aberrations) that the system is diffraction limited. Next, the detector shape transfer function corresponds to the shape and size of the individual detectors.

Finally, the detector sample-and-hold transfer function is one that accounts for an integration of the detector output as it scans through space between two sample points. It is a *sinc* function corresponding to the spatial width between two samples. In the horizontal direction, the sample spacing is 0.08 mrad.

The second set of transfer functions combines to give the temporal postfilter MTFs. The temporal postfilter MTF comprises the detector temporal transfer function, the electronics lowpass filter, the electronics highpass filter, and the boosting filter (there is no boost in this example). Because high-performance electronics are inexpensive relative to the other infrared sensor components, these transfer

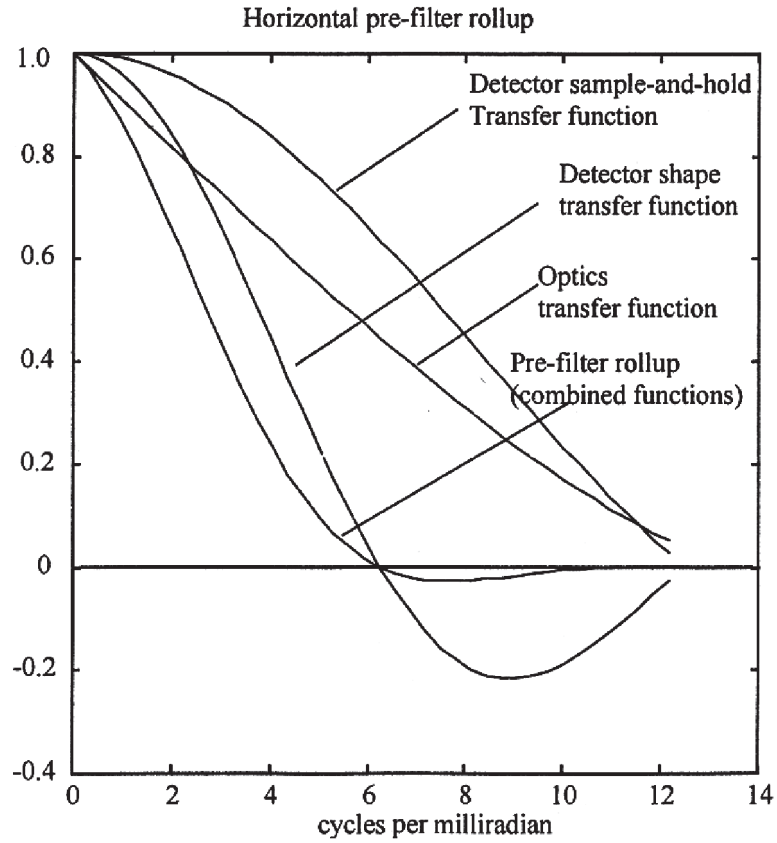


Figure 14.12 Second-generation spatial pre-filter rollup MTF.

functions usually outperform the spatial prefilter and spatial postfilter transfer functions. Because the detector and electronics 3-dB frequencies are given, the temporal postfilter transfer functions can be calculated and are shown in Figure 14.13. The response of the temporal filters can be plotted as a function of temporal frequency (in hertz) or in spatial frequency (cycles per milliradian). The temporal frequency is simply the spatial frequency times the scan velocity. The scan velocity was calculated using (14.50) and (14.51) to be 2,043 mrad/s, and the detector dwell time was calculated to be 0.78 μs.

The spatial postfilter transfer function includes the display spot-size (CRT) transfer function, the display sample-and-hold transfer function, and the eye transfer function. These functions, along with the combined spatial postfilter transfer function, are shown in Figure 14.14.

These transfer functions are described in previous chapters. For the eye and display transfer calculations, the magnification of the system is calculated to be 12.9. The total MTF of the system is the combined prefilter rollup, the temporal postfilter rollup, and the spatial postfilter rollup. These rollups, or combinations of transfer functions, and the system MTF are shown in Figure 14.15.

Now that the system horizontal MTF has been calculated, the spatiotemporal noise is next. We require a system noise bandwidth that is given by (14.52).

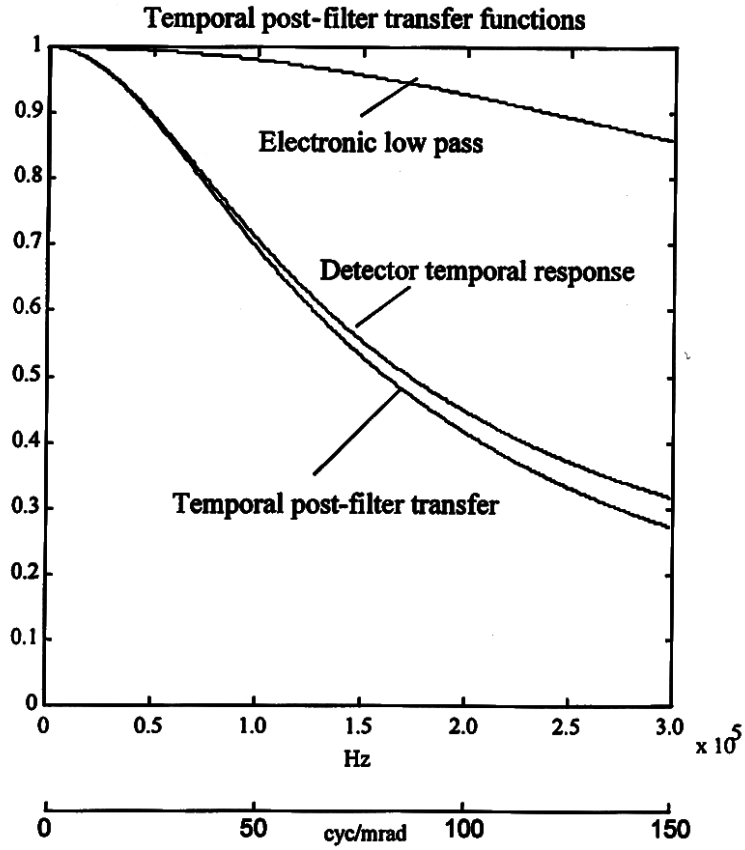


Figure 14.13 Temporal post-filter transfer.

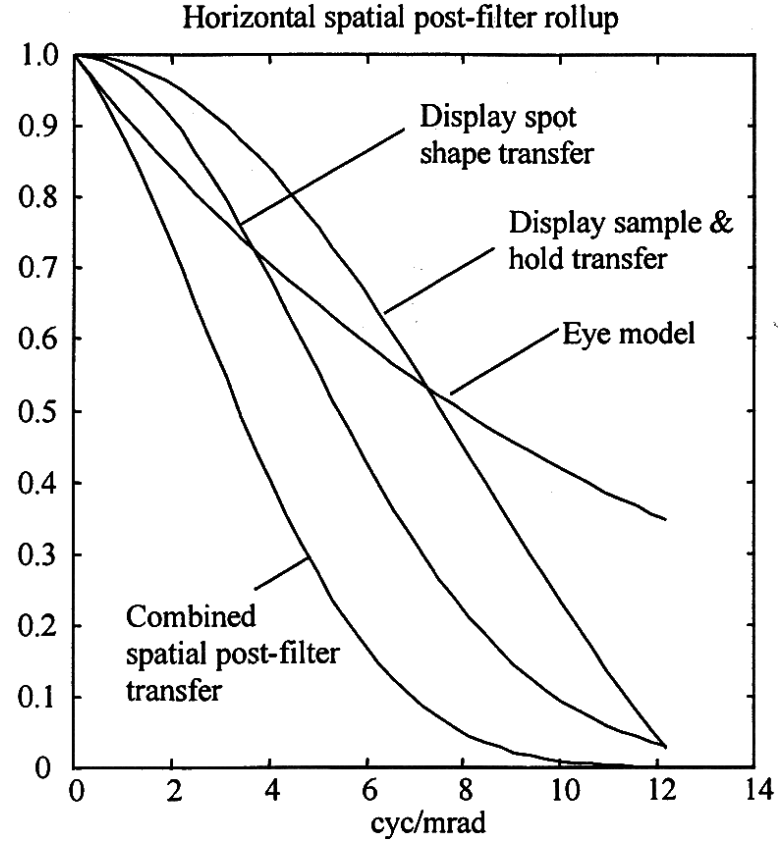


Figure 14.14 Spatial post-filter transfer.

Because this is a second-generation FLIR, the integration time of the detector is included in the calculation. In this case, this *sinc* term dominates the calculation. The integration time is given at 10 μs, so the system noise bandwidth is calculated at 4,190 Hz. The spectral detectivity is the peak detectivity times the normalized detectivity at each of the detectivity wavelengths. This detectivity is used in the denominator integration of (14.45) to give a spatiotemporal noise of 0.062K.

The rest of the MRT parameters are the threshold SNR (given at 2.5), $k_h(\xi)$, and the eye integrals. We set $k_h(\xi)$ to 1 because the 3D noise parameters are not given. Therefore, the system is assumed to be limited only by random spatiotemporal noise and not pattern noise. Alternatively, the estimates in Table 14.4 for typical systems may be used. Finally, the eye integrals are estimated. The temporal integral E_t is simple at 0.333. The two spatial integrals given by (14.48) and (14.49) are provided in Figure 14.16. We now have all the parameters and functions necessary to calculate the horizontal MRT. The result is typically plotted on a semilog scale and is shown in Figure 14.17.

The sample rate in both the vertical and horizontal directions was 12.5 cycle/mrad and the half-sample rate (sometimes called the *Nyquist rate*) is 6.25 cycle/mrad. The MRT plot matches up high-temperature values very well to a spatial

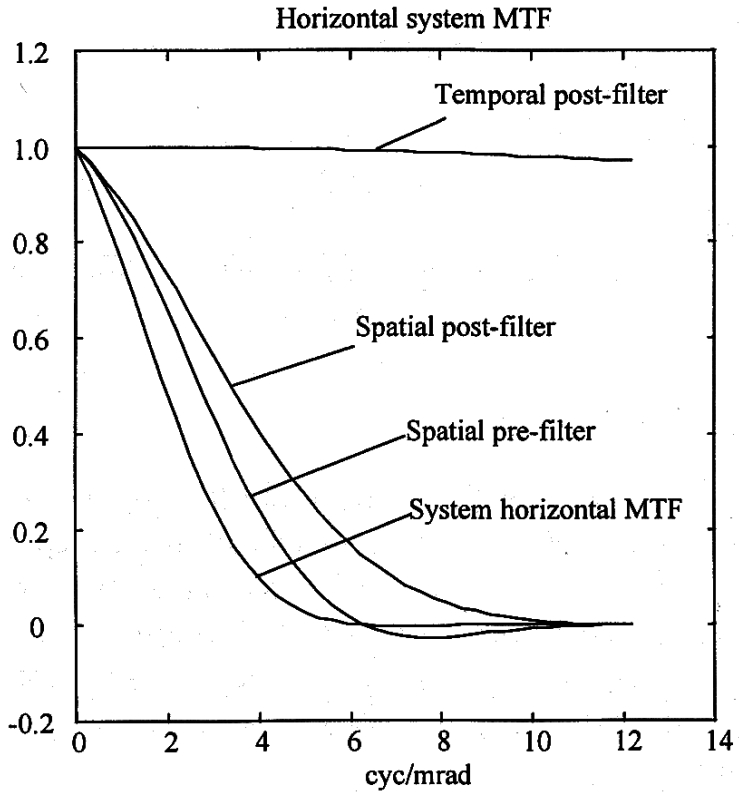


Figure 14.15 System horizontal MTF.

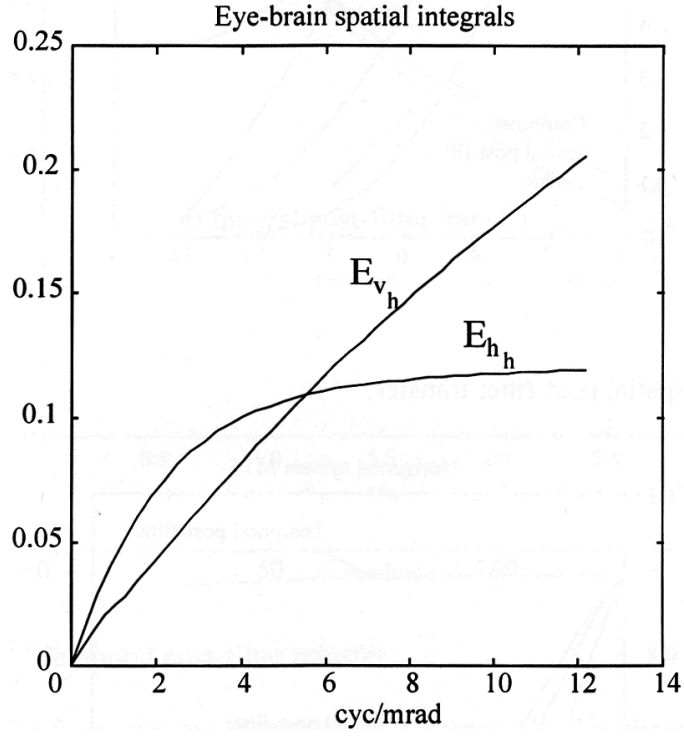


Figure 14.16 Spatial eye integrals.

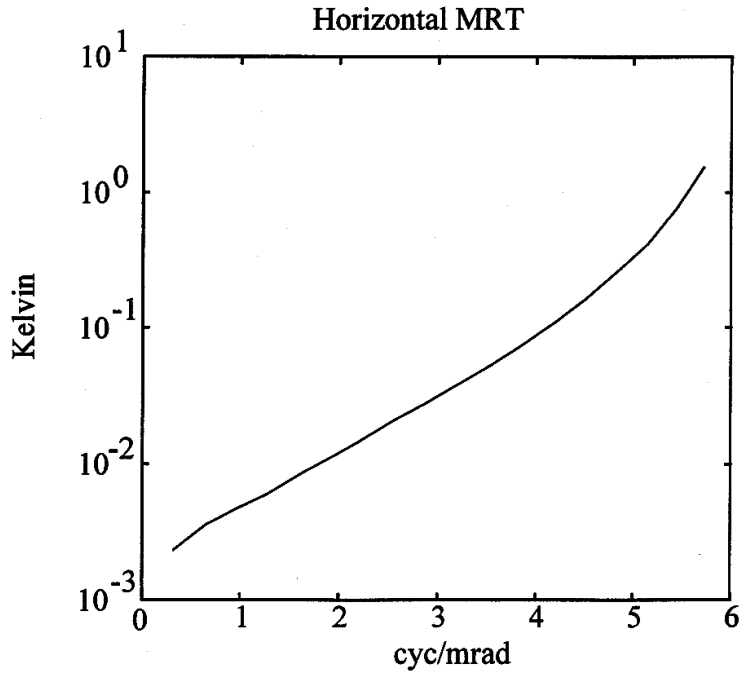


Figure 14.17 Second-generation FLIR example MRT.

frequency near this Nyquist limit. Therefore, the MRT shown is ready to use in acquisition modeling (presented later in this chapter). This is not the case for staring sensors. They are sometimes undersampled significantly.

In the previous version of this book, we provided an MRT example for staring sensors where the frequency limit of the staring sensor was limited by the Nyquist limit. We are not providing a staring array example of MRT in this book since there are many problems with MRT and staring arrays. However, we will just state that MRT has been used on staring arrays with some success, but sampling becomes a problem. For staring arrays (and even scanning systems) it is recommended to use the TTP metric provided later in this chapter.

Two-Dimensional MRT

A 2D MRT is determined using the vertical and horizontal MRTs found in the previous section. Consider the horizontal and vertical MRTs shown in Figure 14.18. The spatial frequencies of the horizontal and vertical MRTs are noted when both of the MRTs are matched. The geometric mean of the horizontal and vertical spatial frequencies gives the 2D MRT spatial frequency

$$\rho_{2D} = \sqrt{\xi\eta} \tag{14.61}$$

The matching MRT is then plotted as a function of the 2D spatial frequency. This new function is the 2D MRT. Note that the conversion is a spatial frequency conversion and no manipulation is performed on the two differential temperatures.

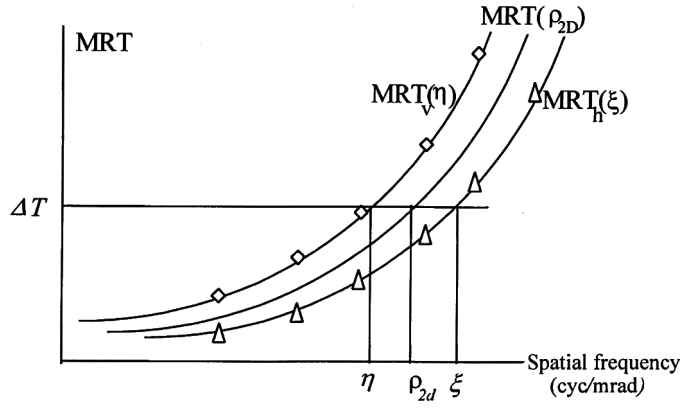


Figure 14.18 Two-dimensional MRT.

14.7 Target Acquisition with Limiting Frequency (Johnson's N_{50})

One of the primary system level tactical sensor performance measures is the probability of performing a visual discrimination task (detection, recognition, and identification) where the sensor is at a given range from a target. Over the years, a variety of researchers [13, 14] have demonstrated that the level of target discrimination possible with a system is related to the resolvability of barchart equivalents of the target. Here, we focus on static performance, which refers to the ability of the observer to perform a discrimination task while having prior knowledge that a target is within the FOV and given an infinite amount of time. The target, scene, and sensor are stationary.

The performance measures of discrimination were first quantified by Johnson [13] in 1958 and have been modified slightly over the years for image intensifiers and infrared sensors since the work was based on image intensifier imagery. As a result of Johnson's original performance measures, the *limiting frequency approach* discrimination criteria (with 50% probabilities) are shown in Table 14.6. Johnson's development was based on using an image intensifier and the U.S. Air Force's tribar chart, where he sought a way to relate in-the-field performance of image intensifiers to objective laboratory measures.

Table 14.6 Discrimination Criteria Using Johnson's Approach (Limiting Frequency)

Discrimination Level	*Meaning*	*1D Cycles Across Minimum Dimension,* N_{50}	*2D Cycles Across Critical Dimension,* N_{50}
Detection	An object of military interest is present (change FOV to narrow to interrogate)	1	0.75
Recognition	Class to which an object belongs (tank, truck, or armored personnel carrier)	4.0	3.0
Identification	Specify type of object (M1A tank, T62 tank, T72 tank, etc.)	8.0	6.0

The initial discrimination criteria developed by Johnson were modified to be relevant for FLIR performance prediction by Ratches et al. [14] in 1975. This performance prediction as a function of range was based on a 1D model using only the horizontal MRT. The technique was acceptable because most FLIRs at the time included a fixed relationship between the vertical and horizontal resolution (approximately 2:1). The work was further modified to reflect the 2D model that is derived using both the horizontal and vertical MRTs. The 50% probability cycle criteria for the 1D and 2D MRTs are shown in Table 14.6.

A number of assumptions have been made over the years in the development of the discrimination performance measure. The first is that the majority of targets that were used in field tests for the validation of the models were either high-contrast or high-differential-temperature targets. Probability of discrimination increases (within limits) as target viewing time, angular subtense, and contrast increase. The work performed to date was based mostly on targets with a maximal height-to-width ratio of 2:1. Other significant work was performed by Moser [15], which dealt with targets that had a larger width-to-height ratio (e.g., ships). It is unclear exactly how medium-size targets, such as rocket launcher platforms, should be modeled. In recent years, much work has been aimed at overcoming the inherent limitations of the NVESD performance model and incorporating the aspects and issues of the second-generation FLIRs and, ultimately, staring array imagers. The result was the target task performance metric and the new method for calculating range performance (presented in the previous chapter and the next section for imagers).

The procedure for producing a probability of detection, recognition, or identification curve using Johnson's approach is quite simple. Consider the process flow given in Figure 14.19. Four parameters are needed to generate a static probability of discrimination curve as a function of range: the estimated target-to-background temperature differential, estimated height and width of a target, estimate for the atmospheric transmission of the band of interest for a number of ranges around the ranges of interest, and the sensor 2D MRT (either modeled or measured on a real system).

The target parameters are determined first. The critical dimension of the target is taken as either the root of the area of the target (in meters) or, if the actual area is not known or measured, the geometric mean of the target height and width:

$$d_c = \sqrt{wh} \quad \text{[m]} \tag{14.62}$$

The target-to-background temperature difference is then estimated based on target and background characteristics. For ground targets, these differential temperatures are usually between 1.25° and 4.0°C. The apparent differential temperature is then determined. Numerous techniques are available for determining the apparent differential temperature and the simplest form is the RSS delta T (from Chapter 5) times the atmospheric transmission.

Once an apparent differential temperature has been obtained, the highest corresponding spatial frequency that can be resolved by the sensor is determined. This is accomplished by finding the spatial frequency ρ_o that matches the target apparent

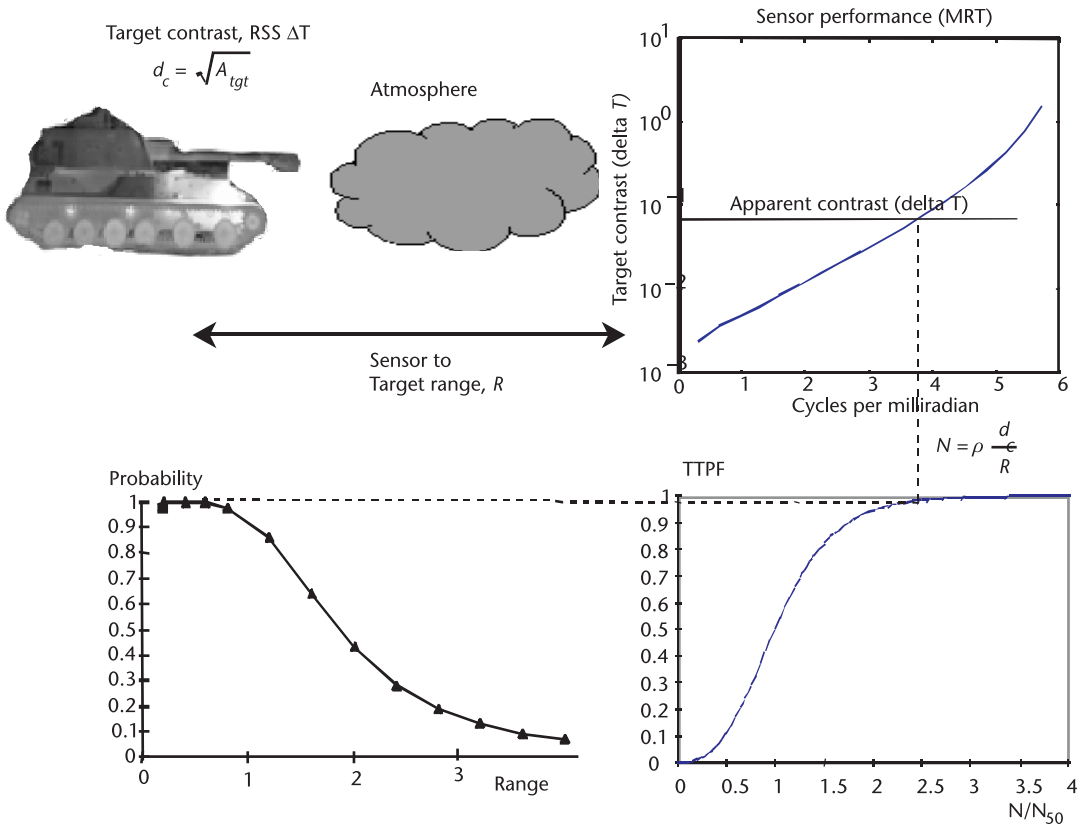


Figure 14.19 Probability of discrimination process.

differential temperature on the MRT curve. The number of cycles across the critical target dimension that can actually be resolved by the sensor is

$$N = \rho_o \frac{d_c}{R} \quad \text{[cycles]} \tag{14.63}$$

where d_c is the critical target dimension in meters, and R is the range from the sensor to the target in kilometers.

The probability of discrimination is determined using the target transfer probability function (TTPF). The level of discrimination (detection, recognition, or identification) is selected from Table 14.6 and the corresponding 50% cycle criteria, N_{50}, is taken from the table. The probability of detection, recognition, or identification is then determined with the TTPF as follows:

$$P(N) = \frac{\left(\frac{N}{N_{50}}\right)^{2.7+0.7\left(\frac{N}{N_{50}}\right)}}{1+\left(\frac{N}{N_{50}}\right)^{2.7+0.7\left(\frac{N}{N_{50}}\right)}} \tag{14.64}$$

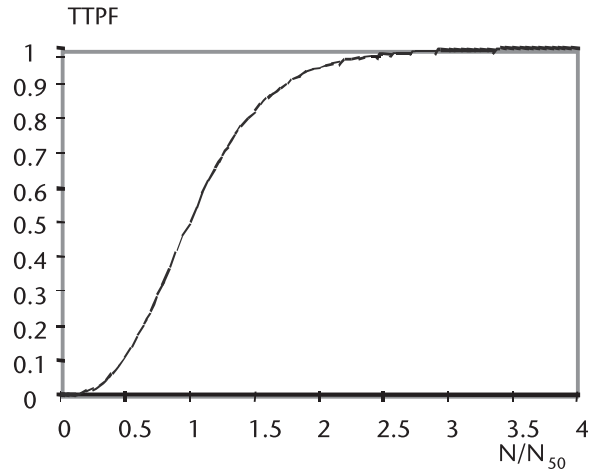


Figure 14.20 TTPF function.

This function is shown in Figure 14.20. The probability of discrimination task is then assigned to the particular range. A typical probability of discrimination curve has the probability plotted as a function of range. Therefore, the above procedure would be repeated for a number of different ranges.

While the following may be obvious, there are a number of characteristics that improve probability of detection, recognition, and identification in imaging systems. Improvements are seen with larger targets, larger target-to-background contrast, larger target emissivities, larger atmospheric transmission, smaller MRT values (as a function of spatial frequency), and usually smaller FOVs (if the target does not have an extremely small differential temperature).

An example of the probability of discrimination procedure is worthwhile because this procedure is one of the more useful ones in sensor analysis. It allows for the conversion of the sensor performance MRT specification into an operation performance specification. In fact, it was used frequently to give the probability of discrimination regardless of the sensor MRT source. The output of a sensor model or an actual measured (laboratory) sensor MRT can be used. Consider the 2D longwave sensor MRT given in Table 14.7.

We want to determine the sensor probability functions of detection, recognition, and identification of a standard NATO target at a range of up to 20 km. A standard NATO target, which is representative of the front aspect view of a tank, has a width of 2.3m, a height of 2.3m, and a target-to-background differential temperature (blackbody equivalent assuming a 300°C background) of 1.25°C. The conditions are clear, the standard U.S. atmosphere, and an altitude of 0.5 km above sea level. MODTRAN has been run for this condition with each of the ranges desired and the longwave transmission (8 to 12 μm) is given in the second column of Table 14.8.

The calculation begins by determining an apparent target-to-background differential temperature as seen by the sensor. In the longwave region, it is common to multiply the actual differential temperature by the atmospheric transmission. This procedure is accompanied by an error of a few percent. If this had been a midwave

Table 14.7 Two-Dimensional MRT for Example Using Johnson's Approach

ρ_{2D} (cycle/mrad)	MRT_{2D} (K)
0.297	0.010
0.599	0.021
1.091	0.043
1.692	0.090
2.265	0.185
2.754	0.382
3.148	0.790
3.440	1.631
3.649	3.369
3.778	6.958

(3- to 5-μm) sensor, the error would have been larger and we would have had to count photons through the atmosphere to obtain a radiometric apparent differential temperature at the imager.

Here, because we are working in the longwave region, we multiply the target-to-background differential temperature by the atmospheric transmission to determine the apparent differential temperature:

$$\Delta T_{app}(R) = \Delta T_{tgt}\,\tau_{atm}(R) \quad [\mathrm{K}] \tag{14.65}$$

The apparent differential temperature is given in the third column of Table 14.8. For each $\Delta T_{app}(R)$, a frequency ρ_o must be determined using the sensor 2D MRT. Recall that ρ_o is the highest spatial frequency that can be resolved by the sensor at a given $\Delta T_{app}(R)$. Interpolation is usually required to give an accurate ρ_o.

Take, for example, R equal to 6 km in the table. The apparent target-to-background temperature is 0.80°C, which lies between the 0.790 and 1.631 MRT points in Table 14.7. The corresponding spatial frequencies are 3.148 and 3.440 cycle/mrad, respectively. Interpolating, the spatial frequency corresponding to 6 km is 3.1 cycle/mrad. Interpolations are calculated for each of the apparent differential temperatures to give the critical frequency as a function of range. These are given in the

Table 14.8 Probability of Discrimination Example

Range (km)	$\tau_{atm}(R)$	$\Delta T_{app}(R)$ (K)	ρ_o (cycle/mrad)	N *Number*	P_{ID}	P_{Rec}	P_{Det}
1	0.9	1.125	3.26	7.5	0.69	0.98	1
2	0.832	1.04	3.23	3.7	0.18	0.68	1
3	0.775	0.969	3.21	2.46	0.07	0.34	1
4	0.725	0.906	3.19	1.83	0.03	0.17	0.98
6	0.638	0.799	3.15	1.2	0.01	0.06	0.86
8	0.566	0.707	3.07	0.88	0	0.03	0.64
10	0.504	0.63	2.99	0.69	0	0.02	0.43
12	0.45	0.562	2.92	0.56	0	0.01	0.28
14	0.403	0.504	2.87	0.47	0	0	0.19
16	0.362	0.452	2.82	0.4	0	0	0.13
18	0.325	0.406	2.77	0.35	0	0	0.09
20	0.292	0.365	2.71	0.31	0	0	0.07

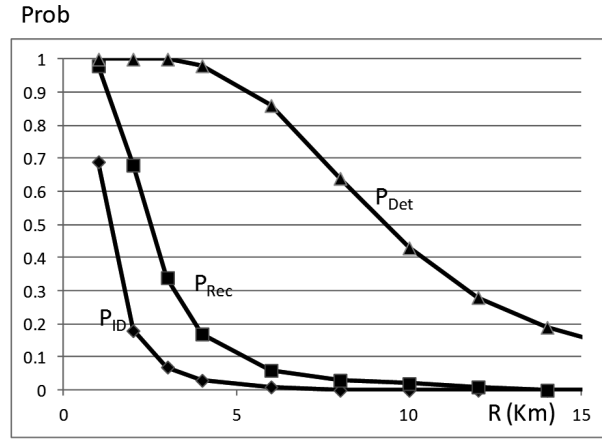


Figure 14.21 Probability of discrimination example.

fourth column of Table 14.8. The critical number of cycles is then calculated using the critical frequency, the range, and the critical target dimension. The geometric mean of the height and width gives a critical dimension of 2.3m. The number of resolvable cycles across the target is given in the fifth column of Table 14.8. Finally, the probabilities of identification, recognition, and detection are given in the sixth, seventh, and eighth columns, respectively. The 50% number of cycles across target N_{50} for identification, recognition, and detection were taken as 0.75, 3.0, and 6.0, respectively. A graphical representation is given in Figure 14.21.

14.8 System Contrast Threshold Function

Johnson's approach (sometimes called ACQUIRE) from the previous section worked well for many years and provided reasonable performance estimates for first- and second-generation thermal imagers. A problem was raised when comparing staring array imagers that were ultimately limited by the half-sample rate (Nyquist rate), but were very different in sensitivity. Two such imagers were PtSi versus InSb, where PtSi would have the same focal plane geometry as that of InSb, but PtSi imagers produced more noise than InSb imagers. Johnson's approach provided the same range performance predictions since the range calculation depended on one "limiting frequency" for each range. This was a real problem since the two imagers would yield different range performance in field trials.

Vollmerhausen et al. [16] developed the TTP metric to include all of the frequencies that were seen by the observer as well as how much more contrast the target provided beyond the observer contrast threshold. The TTP uses the contrast threshold function of the entire imaging system, which is the observer CTF (from Chapter 13) with the added magnification, blur, and noise provided by the imaging system. The system CTF [17, 18] is

$$CTF_{sys}(\xi) = \frac{CTF\left(\dfrac{\xi}{M_{sys}}\right)}{H_{sys}(\xi)}\left[1+\frac{n_{eyes}\gamma^2\sigma^2(\xi/M_{sys})Q_t}{L_o^2+\dfrac{\beta^2}{A}L_oQ\in}\right]^{\frac{1}{2}} \quad \text{[unitless]} \qquad (14.66)$$

where

CTF = contrast threshold function of the human visual system
M_{sys} = system magnification because the CTF is defined in eye space
H_{sys} = system MTF
σ^2 = variance of the external noise 1 sec after it has been filtered by the system and eye MTFs
ξ = horizontal target frequency
L_0 = mean luminance of the display in foot-Lamberts
n_{eyes} = number of eyes
γ= noise calibration constant with a value of 240
Q_t = combined temporal bandwidth of the eye and display
β = eye photon noise calibration constant with a value of 4.9
A = area of the pupil
$Q_\in$ = temporal bandwidth of the eye.

Preece et al. [17] explain that to properly combine external noise with internal eye noise, the noise is added just before the decision threshold at the end of the human visual system (HVS) processing chain. The following discussion is provided by Brad Preece. External noise is added to the CTF of the HVS via a spatial-frequency dependent noise factor. This factor is defined using the ratio of external noise to the internal noise and has the general form

$$CTF_{sys}(\xi) = \frac{CTF\left(\dfrac{\xi}{M_{sys}}\right)}{H_{sys}(\xi)}\left[1 + \frac{\sigma_d^2}{\sigma_\in^2}\right]^{1/2} \tag{14.67}$$

where σ_d^2 is the filtered external noise variance, and $\sigma_\in^2$ is the total effective internal eye noise variance. The external noise is processed by all of the appropriate HVS spatial and temporal filters while the perceived eye photon noise is modified only by the HVS temporal bandwidth. External noise, denoted as N_d^2, is a power spectral density (psd) ($fL^2 \cdot mrad^2 \cdot s$) of the luminance noise seen by the observer. The external noise is filtered by postfilter MTFs (H_{post}), which include all MTFs after the detector sampling function, the eye MTF (H_{eye}), and an eye channel MTF (H_{chan}). The eye channel MTF used was proposed by Barten [19] to model the noise-masking effects of the HVS. The external noise is filtered temporally by the combined temporal bandwidth of the display and the HVS, denoted as Q_t. The number of eyes does not affect display noise because the noise is correlated for both eyes. The luminance noise pdf becomes a unitless contrast by dividing by the mean luminance squared L_0^2. The complete expression of unitless contrast variance for external noise is

$$\sigma_d^1(\xi, L_0) = \frac{N_d^2}{L_0^2} Q(\xi) Q_t \tag{14.68}$$

where $Q(\xi)$ is the spatial bandwidth in cycles per milliradian to the eye and Q_t is the temporal bandwidth in Hertz. Each of these bandwidths is defined below.

The eye photon noise variance is proportional to the illuminance at the retina, relating directly to the area of the pupil. Other unknown scaling factors related to photon collection and bandwidth normalization are included in a constant κ. Although it is reasonable to assume the eye photon PSD is filtered by the eye channel MTF, H_{chan}, measured data do not agree with that assumption. The effective temporal bandwidth of the HVS, denoted as Q_ϵ, is included. The complete expression of unitless contrast variance for eye photon noise is

$$\sigma_p^2(L_0) = \frac{\kappa^2}{AL_0} Q_\epsilon \tag{14.69}$$

The last internal noise expression is an unknown eye–brain noise floor and this unitless contrast variance is denoted as σ_B^2. The internal noise benefits from the number of eyes used and can be expressed as the sum of the two internal eye variances:

$$\sigma_\epsilon^2(L_0) = \frac{1}{n_{eyes}}\left[\sigma_B^2 + \sigma_p^2(L_0)\right] \tag{14.70}$$

For easier comparison the constants σ_B and κ are replaced with model calibration constants, γ and β using the following definitions:

$$\gamma = \frac{1}{\sigma_B} \text{ and } \beta = \kappa\sigma_B \tag{14.71}$$

Each of the spatial and temporal bandwidths used above is given by

$$Q(\xi) = \int\int_{-\infty}^{\infty} \left|H_{post}(\xi',\eta')H_{eye}(\xi',\eta')H_{chan}(\xi',\xi)\right|^2 d\xi'\, d\eta' \tag{14.72}$$

$$Q_t = 2\int_0^{\infty} \left|H_{disp}(f,\tau_D)H_t(f,\tau_{eye})\right|^2 df \tag{14.73}$$

$$Q_\epsilon = 2\int_0^{\infty} \left|H_t(f,\tau_{eye})\right|^2 df = \frac{1}{\tau_{eye}} \tag{14.74}$$

where η' and ξ' are orthogonal noise frequencies, f is the temporal frequency in hertz, and H_{disp} is the temporal MTF of the display with a frame time of τ_D. The temporal eye filter is modeled as a first-order exponential giving

$$\left|H_t(f,\tau_{eye})\right| = \frac{2/\tau_{eye}}{\sqrt{(2/\tau_{eye})^2 + (2\cdot\pi\cdot f)^2}} \tag{14.75}$$

where τ_{eye} is the effective integration time of the HSV. The channel filter is defined as

$$H_{chan}(\xi',\xi) = e^{-2.2\left[ln\left(\frac{\xi'}{\xi}\right)\right]^2} \tag{14.76}$$

The final expressions for σ_d^2 and σ_ϵ^2 are expressed as

$$\sigma_d^2(\xi, L_0) = \frac{N_d^2 Q(\xi)}{L_0^2} Q_t = \frac{\sigma^2(\xi)}{L_0^2} Q_t \tag{14.77}$$

$$\sigma_{\epsilon}^2(L_0) = \frac{1}{n_{eyes}}\left[\frac{1}{\gamma^2} + \frac{\beta^2}{\gamma^2 A L_0}Q_{\epsilon}\right] \tag{14.78}$$

Equation (14.66) has been simplified using the following

$$\sigma^2(\xi) = N_d^2 Q(\xi) \tag{14.79}$$

The pupil diameter and eye integration time are given in Chapter 13. Note from (14.66) that if the imager noise is small, then the system CTF is simply the human eye CTF modified by the magnification and the system MTF.

The system CTF described in this section is appropriate for most imagers including visible, image intensifiers, SWIR, MWIR, and LWIR sensors. The parameters are provided such that they can be calculated for any of these systems.

Blur and noise are covered in the system CTF and the overall performance of the system so far. However, to include sampling artifacts in the range performance calculations, one final modification must be implemented. This modification is implemented on the TTP metric calculation provided in the next section.

Figure 14.22 is three panels prepared by Eddie Jacobs at the University of Memphis and it demonstrates the characteristics of a system CTF. The panel on the left shows how the eye CTF works where the contrast of the sinusoids decreases from top to bottom. The eye CTF shows that midfrequencies can be seen very well and the lower and higher frequencies are only seen at higher contrasts. The resolved cycles follow a CTF curve. Then a system MTF (where the system has very low noise) blurs the image, as seen in the middle panel. The higher frequencies that are blurred by the system are no longer seen by the observer. Finally, in the final and right panel, system noise is added which makes the lower contrast frequencies harder to see.

Example 14.5

The goal of this example is to demonstrate the CTF noise model of (14.66) with a real system as described in Table 14.2. The system has a NETD of 15 mK; however, we vary the amount of NETD in order to show how the system CTF behaves as the imager becomes noise limited. This is accomplished by keeping the amount of signal constant and adding a secondary noise to the system such as readout. The system provides a signal of 1.512×10^7 electrons collected per pixel per frame at a target temperature of 300K. The background is assumed to be at 299K, giving a signal of 1.47×10^7 electrons per pixel per frame and an RSS temperature difference of 1K.

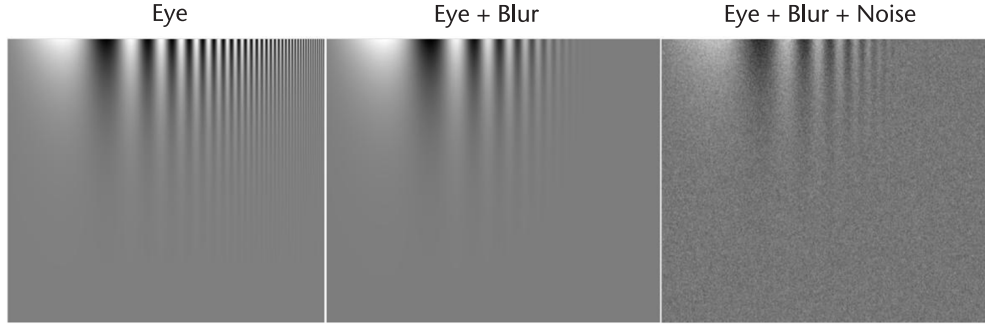


Figure 14.22 System contrast threshold function demonstrated. (Courtesy of Eddie Jacobs.)

Table 14.9 CTF Noise Example Values at 10 fL

G_{det}	422,200 at 300K target and 1K RSS
G_{disp}	9.35e–6 at 10 ftL
σ_{ftL}	0.0592 at 15 mK
N_d^2	2.597e–5 at 15 mK
Q_t	3.77 at 60FPS
$Q_\in$	4.0323 at 10 ftL
τ_{eye}	0.2480 at 10 ftL
Q_t	3.77 at 10 ftL and 60 frame/s
n_{eyes}	2
A	10.2611 at 10 ftL

Under these conditions and neglecting atmospheric loss, the RMS noise for 15 mK is 6,333 electrons per pixel per frame. The ratio between NETD and electrons is defined as the detector noise gain (G_{det}). Calculated values are provided in Table 14.9. Once the pedestal between the target and background is subtracted in order to display the image, a second gain can be defined (G_{disp}) between the number of electrons or counts in the system and the average luminance on the display. The displayed noise variance due to a NETD is then calculated as follows:

$$\sigma_{ftL}^2[ftL]^2 = NETD^2[K]^2 \cdot G_{det}^2 \left[\frac{(e^-/\text{pixel frame})}{k}\right]^2 \cdot G_{disp}^2 \left[\frac{f+L}{(e^-/\text{pixel frame})}\right]^2 \tag{14.80}$$

A psd for a white noise floor is defined as the variance multiplied by its bandwidth. The CTF noise model requires a psd for displayed external noise (N_d^2) with units of $fL^2 \cdot mrad^2 \cdot s$ and is solved for by

$$N_d^2[fl^2 \cdot mrad \cdot \text{s}] = \sigma_{ftL}^2[ftL]^2 \cdot \frac{\Delta p_{disp}^2}{VD^2}\left[\frac{mm}{m}\right]^2 \cdot \Delta\tau[sec] \tag{14.81}$$

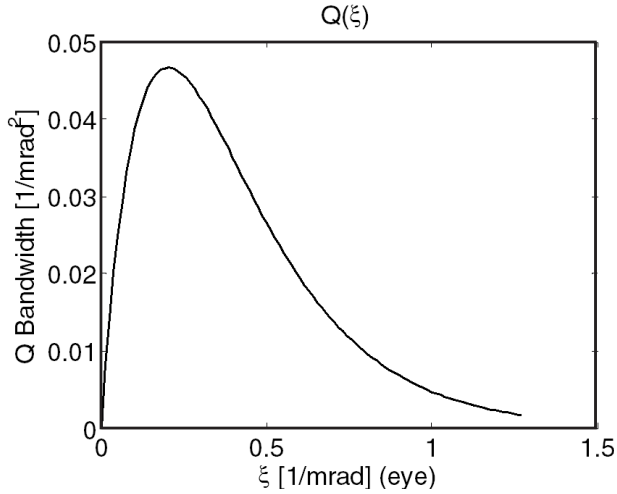


Figure 14.23 Q bandwidth calculated from (14.72). The MTFs used for the calculations include the display *sinc* MTF, the eye MTF, and the channel MTF.

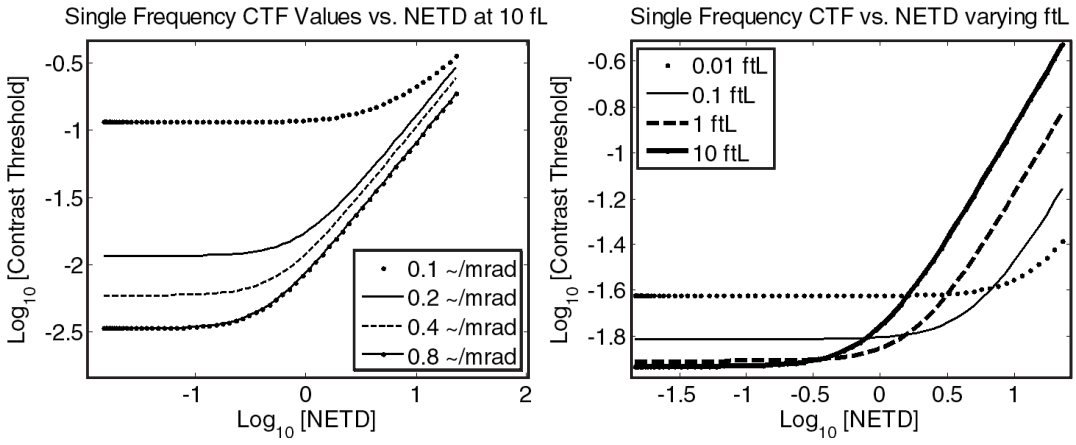


Figure 14.24 (a) The system contrast threshold function plotted at four different frequencies vs. NETD. (b) Single frequency (0.2 eye/mrad) plotted at different luminance values versus NETD.

where Δp_{disp} is the display pixel pitch and VD is the viewing distance. Note that the ratio of pixel pitch in millimeters over the viewing distance in meters gives the angle in millliradians at the eye using the small angel approximation. The display used for this example is an LCD with a pixel pitch of 0.2 mm and a viewing distance of 30 cm. How to calculate the system magnification (approximately 20) is left to the reader. Note that the example calculations are done in eye space, but need to be scaled to object space for the TTP metric defined in the next subsection. To calculate Q(ξ) [see (14.72)] H_{post} is a *sinc* MTF from the rectangular reconstruction filter.

To calculate the total noise variance seen by the observer, the PSD (N_d^2) is multiplied by the Q in Figure 14.23. Now that all parameters have been calculated, the single frequency values of the system CTF are plotted vs. NETD (Figure 14.24). The CTF can be thought of as the noise floor of the human eye. A lower system CTF represents a more sensitive system. When a system has a small NETD, the noise is below the human threshold and internal eye noise dominates. As external noise increases (sensitivity of the system decreases), it starts to overcome and dominate internal eye noise. The knee in the curve represents the point at which external system noise is equal to internal eye noise.

When lowering the luminance on the display, photon noise inside the eye dominates, causing the CTF to increase. Therefore it takes more external noise at lower luminance to affect performance. Similar graphs can be made varying the frame rate on the display and showing the difference between monocular and binocular vision.

14.9 Target Acquisition with the Target Task Performance Metric (and Vollmerhausen's V_{50})

Now that we have a $CTF_{sys}(\xi)$, the method for calculating range performance is a little different than that of the limiting frequency approach using the MRT or MRC. The TTP metric is the integral of the root of the apparent target contrast over the system CTF:

$$TTP = \int_0^{\xi_{cut}} \sqrt{\frac{C_{tgt}}{CTF_{sys}(\xi)}} d\xi \tag{14.82}$$

where C_{tgt} is the apparent target contrast after taking into consideration the atmosphere and ξ_{cut} is the frequency where the apparent target contrast is equal to the system CTF. Note that the integral includes all frequency content that can be viewed by the observer.

In the previous section, we stated that aliasing is not included in the system CTF. It turns out that in-band aliasing does not degrade performance significantly, but the display artifacts seen as "out-of-band aliasing" or out-of-band spurious response do degrade performance. The calculation of out-of-band spurious response is provided in Chapter 3, but the postfilter MTF must include the eye MTF. The spurious response ratio (SRR) of the out-of-band spurious response integral to the baseband integral [e.g., (3.49)] can be found:

$$SRR = \frac{SR_{OB}}{R_{BB}} \tag{14.83}$$

A reduction in the TTP metric value due to out-of-band aliasing is applied as:

$$TTP_{sr} = TTP\sqrt{1 - 0.58\,SRR} \tag{14.84}$$

This TTP value now includes the human CTF, system blur, system noise, and aliasing. Preece et al. [18] have investigated out-of-band aliasing as blur and as noise and have considered it as an additional blur and as additive noise.

Similar to Johnson's limiting frequency (or ACQUIRE) approach, the TTP value is converted to a range-dependent parameter of integrated cycles on target by

$$V(R) = \frac{L_{tgt}}{R} TTP \tag{14.85}$$

where L_{tgt} is the characteristic dimension of the target and R is the range from the imager to the target. The effective cycles on target, V, are then used to predict the probability of task performance using the logistic function

$$P(R) = \left(\frac{V(R)}{V_{50}}\right)^{1.5} \Bigg/ \left(1 + \frac{V(R)}{V_{50}}\right)^{1.5} \tag{14.86}$$

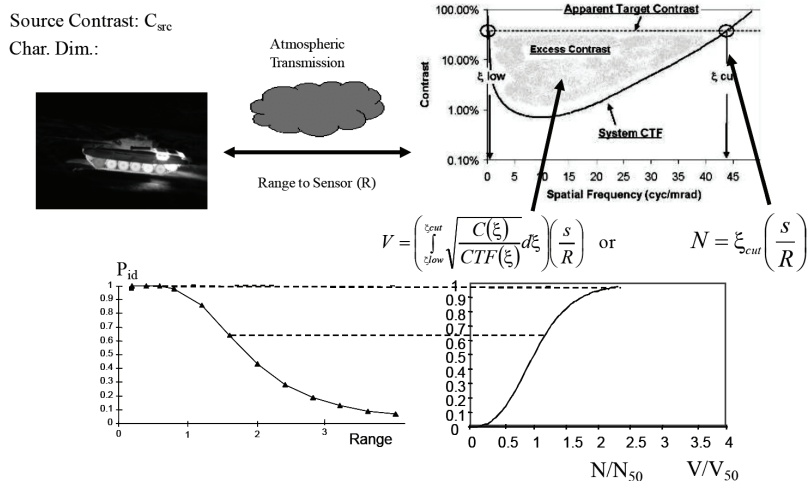


Figure 14.25 Target acquisition approach with the TTP and Johnson approaches.

where V_{50} is the task difficulty parameter. Note that the exponents on the target transfer probability function are different than those used with the Johnson approach. This difference was implemented to update more realistic field measurement conditions. A summary and comparison of the two target acquisition approaches are provided in Figure 14.25.

14.10 Target Sets

Now that we have an approach for calculating the range performance of various imagers, we need the calibration constants for the target sets. Table 14.10 provides the target set parameters for a number of target sets [20]. Moyer et al. provide the table shown giving the typical tank identification criteria along with a number of human target acquisition tasks. The first column shows the discrimination task. The second column shows the spectral band of which the task difficulty has been measured. Each of the task difficulties has been measured with experimentation and the rest of the table provides the results of the experimentation. The $\Delta T/\Delta C$ column provides the appropriate contrast that represents the target set for the particular band. The size in meters provides the square root of the average target area. The N_{50} and V_{50} columns provide the discrimination criteria for the Johnson's approach and for the TTP metric approach, respectively. At the time of this book, these values were tentative and can change based on improvements in the models. For now, they are reasonable estimates for the system performance calculations provided in this book. They are also useful for comparing the difficulty level in discriminating target sets.

The human activity discrimination provided is the ability of a human to determine what activity is being conducted by a person. It was a 24 alternative-forced-choice 24AFC experiment of hostile and nonhostile activities. Figure 14.26 shows the activities used in the experiment and the images shown are in the LWIR. MWIR and LWIR provide similar results.

The weapon/nonweapon target set used included 16 targets all of like size with 8 weapons and 8 objects that were similar size, but were not weapons. The observer had to decide whether the subject was holding a weapon or a nonweapon. The weapon objects varied from individual threats (e.g., AK-47 or M-4 rifles) to squad and vehicle threats (e.g., RPG or SAW). The nonweapon dataset included a tripod, shovel, crowbar, pipe, sledgehammer, rake, two-by-four, and an umbrella.

The two-hand object identification task was to determine the object in a human's hands. For this task, it was no longer sufficient for the observers to simply determine whether the subject was holding a weapon or a nonweapon. In this case they were required to identify the specific weapon or nonweapon that the target was holding. Two different experiments were conducted in order to generate the task difficulty and target set parameters for this task. The target set for this experiment consisted of 10 objects. There was a similar breakdown of weapons versus nonweapons in the set with 5 weapons and 5 nonweapons.

Both the weapon/nonweapon discrimination and the two-handed object identification tasks were conducted for the front aspect and the average of all aspects

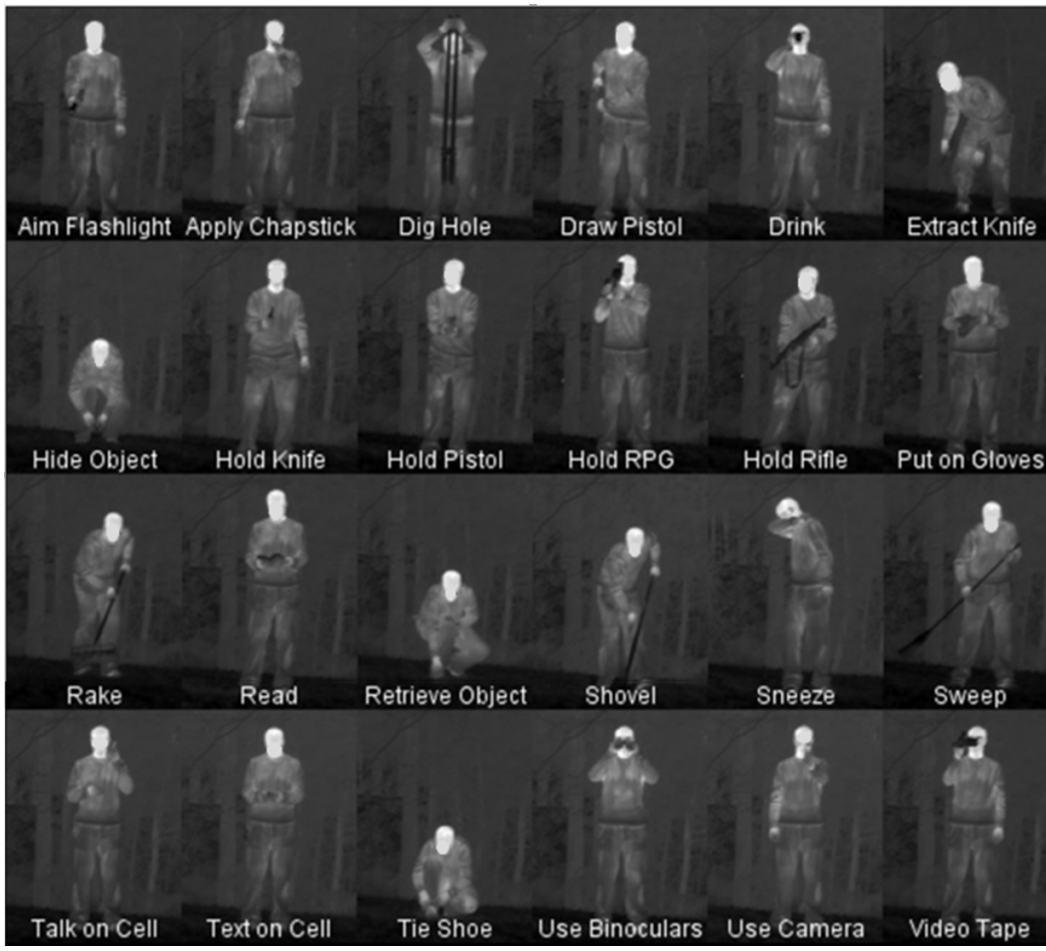


Figure 14.26 Human activity discrimination in the MWIR/LWIR.

(except with the human's back to the imager). Table 14.10 gives results for both cases.

During the past several years, NVESD conducted many studies dealing with various targets [20]. Experiments covered a variety of operating environments including urban, rural, and maritime settings. Experiments also covered a wide variety of tasks, including rural and urban search, swimmer detection, human activity discrimination, classification based on variations in clothing and armament, identification of small handheld and large two-hand objects, and facial identification.

In a personal conversation with Dr. Steve Moyer of NVESD, he stated that the V_{50} and N_{50} terms are model dependent. Any changes to the underlying calculation may result in a change to the V_{50} and N_{50} terms. Since the original Vollmerhausen TTP metric work did not include the strong dependence of human vision CTF to the angle of the display, the V_{50} and N_{50} terms provided in this section are different than those previously described in prior model experiments. It is important to

Table 14.10 Target Set Parameters

Experiment	*Band*	$\Delta T/\Delta C$	*Size* (*m*)	V_{50}	Size/V_{50}	N_{50}	Resolution 90%(cm)
12 target tank ID	Thermal	4	3.11	13	23.9	5	8.5
Human activity (average front aspects)	Passive SWIR	0.25	0.75	6.4	11.7	2.9	3.75
	Visible	0.25	0.75	5.3	14.2	3.0	3.5
	Thermal	3	0.75	8.0	9.4	3.0	3.5
Weapon/nonweapon (average front aspects)	Passive SWIR	0.25	0.25	1.6	15.6	0.8	4.3
	Visible	0.25	0.25	1.6	15.6	0.8	4.3
	Thermal	2	0.25	4.1	6.1	2.1	1.7
Weapon/nonweapon (average all aspects)	Passive SWIR	0.25	0.25	3.4	7.4	1.8	1.9
	Visible	0.25	0.25	×	×	×	×
	Thermal	2	0.25	13.6	1.8	5.9	0.6
Two-hand object ID (average front aspects)	Active SWIR	0.25	0.25	3.6	6.9	2.0	1.4
	Passive SWIR	0.25	0.25	×	×	×	×
	Visible	0.25	0.25	5.3	4.7	2.6	1.4
	Thermal	2	0.25	9.2	2.7	3.7	1
Two-hand object ID (average all aspects)	Active SWIR	0.25	0.25	15.3	1.6	5.7	0.5
	Passive SWIR	0.25	0.25	10.3	2.4	4.6	0.75
	Visible	0.25	0.25	×	×	×	×
	Thermal	2	0.25	19.3	1.3	6.7	0.6

ensure that the task difficulty parameters presented in this paper are used in conjunction with the latest version of imager performance models. The two notable changes introduced since the last release of the NVESD models are the dependence of the observer CTF on the angle subtended by the target [21] and the introduction of the new noise model, which allows the value of α to vary as a function of luminance [17]. The target parameters provided in this section are preliminary, but are valid for the model approach described in this chapter. All of the previous target experiments are limited to use with the model that was used to determine the target parameters and the calibrations are only useful in those previous models. Rather than provide a large number of model iterations and the associated target iterations, we chose to provide the latest model and target calibration at the time this book was prepared.

14.11 Classic ISR, NIIRS, and General Image Quality

As reconnaissance systems evolved in the 1970s, physical measures such as scale and resolution no longer adequately defined the performance of imaging systems [22]. A perceptually based measure of interpretability was developed called the National Imagery Interpretability Rating Scale (NIIRS). The initial NIIRS was developed for use with visible spectrum imagery. The first version of NIIRS was published as a Standard NATO Agreement but was called the Imagery Interpretability Rating Scale (IIRS). The IIRS is no longer considered valid and has been replaced by an updated version of NIIRS [23]. A separate NIIRS has been developed for use with MWIR systems.

The general image quality equation (GIQE) was developed to provide NIIRS with predictions for a given system design and operating parameters. Version 4.0 [24] was released in 1996. The GIQE is currently being used to model unmanned aerial vehicle (UAV) sensor performance as well as upcoming commercial satellite systems. Performance goals for the Predator and Global Hawk UAVs have been specified in terms of the NIIRS. It is expected that this model will become the standard in the acquisition process for the evaluation of ISR sensors. Also, as ISR sensors become more capable and can support tactical missions, it is important to understand the relationship between this ISR model and the tactical MRT acquisition model.

National Imagery Interpretability Rating System

Before considering the GIQE, the reader must understand the NIIRS scale. The NIIRS scale is an intelligence community measure that is used by imagery analysts (IAs) to perform quantitative judgments on the interpretability of a particular image. The process of rating an image corresponds to the assignment of a number that describes the image interpretability. The NIIRS rating has been developed to ensure that imagery collection and exploitation meet the informational needs of the end user. The NIIRS is a scale from 0 to 9 that can be used to compare the products of different imaging systems. The scale has become an important tool for defining image requirements, selecting and tasking imaging systems, providing quality control feedback to operational systems, and specifying performance of new imaging systems.

Table 14.11 Sample NIIRS Descriptions

Rating Level 0	**Rating Level 5**
Interpretability of the imagery is precluded by obscuration, degradation, or poor resolution	Identify Individual Rail Cars by type Identify Radar as Vehicle or Trailer Mounted Identify Tactical Surface to Surface Missile Systems by type
Rating Level 1	**Rating Level 6**
Detect Medium Size Port Facility Distinguish Between Runways and Taxiways	Distinguish Between Models of Small Helicopters Identify Automobiles as Sedans or Station Wagons Identify the Spare Tire on a Medium-Sized Truck
Rating Level 2	**Rating Level 7**
Detect Large Hangars at Airfields Detect Large Static Radars Detect Large Buildings	Identify Individual Rail Ties Identify Fairings and Fitments on Fighter Aircraft Identify Ports, Ladders, Vents on Electronic Vans
Rating Level 3	**Rating Level 8**
Identify the Wing Configuration of Large Aircraft Identify a Large Ship in Port by Type (Cruiser, etc.) Detect Trains or Strings of Standard Rolling Stock	Identify Rivet Lines on Bomber Aircraft Identify a Hand-held Surface-to-Air Missile Identify Windshield Wipers on a Vehicle
Rating Level 4	**Rating Level 9**
Identify All Large Fighter Aircraft Identify, by general type, Tracked Vehicles Detect an Open Missile Silo Door	Identify Truck Vehicle Registration Number Identify Screws and Bolts on Missile Components Detect Individual Spikes in Railroad Ties



Figure 14.27 An NIIRS-4 rated image. (Courtesy of J. Leachtenauer.)

The NIIRS is defined by a set of exploitation tasks or criteria equally spaced across a psychophysically defined quality or interpretability scale [25, 26]. The NIIRS rating of an image defines the exploitation tasks that can be performed on that image. Given a NIIRS rating, all lower valued NIIRS tasks can also be performed on the image. A NIIRS rating accounts for all of the factors affecting interpretability including resolution and sensitivity. Table 14.11 shows a subset of visible NIIRS criteria; a more complete set of all bands has been published in [27]. Note that the NIIRS rating provides inherent target sizing within the definitions. For example, NIIRS 4 allows for identification of large fighter aircraft and NIIRS 6 allows for model discrimination among small and medium helicopters. An example of an image rated at NIIRS 4 is provided in Figure 14.27.

GIQE Model

The GIQE predicts NIIRS as a function of image scale, sharpness or resolution, and SNR. The model was developed based on a large database of images and image analyst (IA) responses and predicts a NIIRS rating with a standard error of 0.3. The GIQE allows system designers and operators to perform trade-offs for the optimization of image quality. The GIQE was originally released in 1994 and has been updated.

The electro-optical GIQE model comprises a constant and four terms:

$$NIIRS = 10.251 - a\log_{10}(GSD_{GM}) + b\log_{10}(RER_{GM}) - 0.656H_{GM} - 0.344(G/SNR) \tag{14.87}$$

where RER_{GM} is the relative edge response (the GM subscript corresponds to the geometric mean of the horizontal and vertical components), GSD_{GM} is the ground

sample distance in inches, H_{GM} is the overshoot from edge sharpening, G is the noise gain from edge sharpening, and SNR is the signal-to-noise ratio. The variables a and b are defined as follows:

$$a = 3.32 \text{ and } b = 1.559 \quad \text{for } RER_{GM} \geq 0.9$$

$$a = 3.16 \text{ and } b = 2.817 \quad \text{for } RER_{GM} < 0.9$$

The relative edge response and overshoot terms require the system edge response. This edge response is found using the system MTF:

$$ER_x(x) = 0.5 + \frac{1}{\pi}\int_0^{\infty} \frac{MTF_x(\xi)}{\xi} sin(2\pi x\xi)\, d\xi \tag{14.88}$$

where x is the response position (in pixels) from the center of a horizontal pixel edge and ξ is the spatial frequency in cycles per sample spacing. This edge response has to be normalized such that when x is a large negative value, ER is 0; and when x is a large positive value, ER is 1 as shown in Figure 14.28. The $ER_y(y)$ calculation is identical, but in the vertical direction. The edge response can be measured in the laboratory with a knife-edge target. The geometric mean of the relative edge response is determined from the product of the differences in edge response at x +0.5 and –0.5 pixels from the edge in each direction:

$$RER_{GM} = \sqrt{[ER_x(0.5) - ER_x(-0.5)][ER_y(0.5) - ER_y(-0.5)]} \tag{14.89}$$

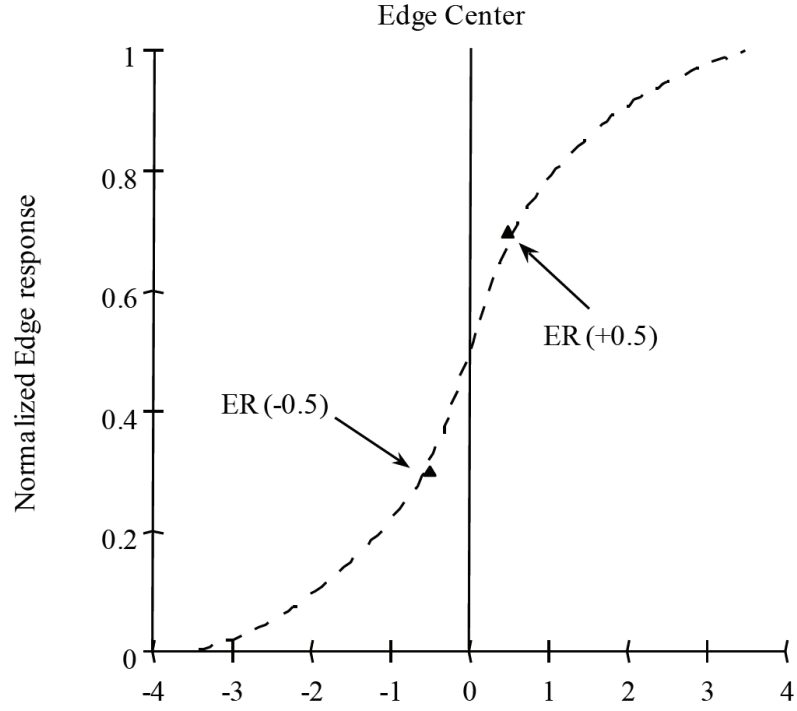


Figure 14.28 Relative edge response.

The ground-sampled distance is the projection of the detector pixel pitch onto the ground, but perpendicular to the sensor's line of sight. The GSD_x term is found by

$$GSD_x = \left[\frac{PixelPitch}{FocalLength}\right]\frac{Range}{cos(\theta_{look})} \quad \text{[in.]} \tag{14.90}$$

where the pixel pitch and focal length are in the same units and the range from the sensor to the target is in inches. The look angle, θ_{look}, is 0 deg for sensors pointing directly downward at the top of targets and nearly 90 deg for targets on the horizon. The geometric mean must be taken of the horizontal and vertical ground sample distances. With EO systems, this term can account for the majority of the NIIRS rating.

The edge height overshoot term is due to *modulation transfer function compensation* (MTFC), or edge sharpening. The term H_x represents the value of the horizontal maximum edge response in the region of 1.0 to 3.0 pixels from the edge [using (14.88)] unless the edge is monotonically increasing. For the monotonically increasing H_x, the edge response at 1.25 pixels from the edge is taken. Again, the geometric mean of the horizontal and vertical responses is used in the equation.

The noise gain G is determined by the gain of the MTFC filter kernel elements:

$$G = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(Kernel_{ij})^2} \tag{14.91}$$

where the MTFC kernel elements are in an $M \times N$ matrix.

The SNR is the ratio of the dc differential scene radiance to the equivalent rms noise radiance. The SNR is the electronic output SNR and does not include MTFC compensation. In EO sensors, the signal is calculated using the reflective properties of the target and background, the atmospheric and illumination conditions, and the sensor characteristics. The noise term includes all of the noise contributors.

This GIQE has been frequently applied to photographic and EO visible systems. It has also been applied to MWIR systems and is currently the infrared design tool for a number of high-altitude UAVs. For the infrared version, the only difference is that the constant in (14.87) becomes 9.751.

Example 14.6

As an example, the GIQE was applied to a narrow FOV EO system. The system FOV was 1.0×1.2 deg with a spectral bandpass of 0.65 to 0.90 μm. The f-number of the sensor was 6.1 and the optical transmittance was 0.91. The focal length was 167 cm and the entrance pupil diameter was 11.3 cm. A 480×640 CCD detector array was used with 15- × 15-μm elements. The center-to-center spacing was 21.8 μrad for a 67% fill factor. The optics were not diffraction limited and an estimate for the optical MTF was used. The display was a 10-fL CRT with a 15.24-cm display height and a 30-cm viewing distance. The eye integration time was set to 0.1 sec and the eye signal-to-noise threshold was set to 2.5. The GSD term provided a significant decrease in the NIIRS estimate with range. The RER term gave a normalized slope of 0.542 to yield an overall decrease of 0.2 NIIRS. The H term decreased the NIIRS by around 0.5. The NIIRS as a function of range is given in Figure 14.29.

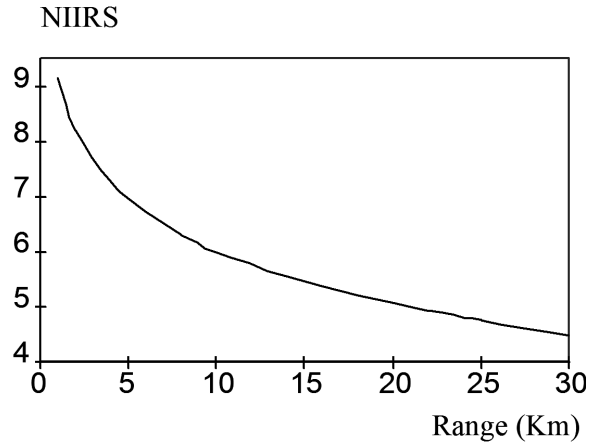


Figure 14.29 GIQE results.

14.12 The Newest Military Imaging Mode: Persistent Surveillance

A new type of military mission and corresponding sensor has been developed and fielded in the past few years. The application is called persistent surveillance (sometimes called PISR) and the sensor is a wide-area coverage (square kilometers) airborne sensor, that provides constant surveillance of a designated region. The region can be an entire town or city and the area coverage by the sensor is relatively large, as shown in Figure 14.30. The sensor does not provide full-motion video as we know it, but it does take a "picture" a few times a second (a rate known as the revisit rate) so as to track objects on the ground. PISR sensors can be visible or SWIR for daytime sensing or infrared for night sensing.

Two common mission requirements for such sensors are the ability to track humans and the ability to track vehicles. Backtracking is similar to tracking, but it is more of a forensic mode. Once a military event of interest occurs (such as a roadside bomb explosion, suicide bomber explosion, or a vehicle bomb explosion), a human or a vehicle is "backtracked" until a point of origin is determined. This point of origin is one that can be interrogated after the fact as a forensic step.

The design goal for persistent surveillance sensors is to maximize wide-area coverage while maintaining enough resolution to perform the mission. The sensors are deployed on balloons at altitudes of 5,000 ft and are also being deployed on aircraft at altitudes of up to 18,000 ft.

Currently, there is no standard model for determining the performance of a sensor design or existing sensor configuration for persistent surveillance. Two primary design parameters are the sensor's ground sample distance (GSD) in meters and the revisit time in frames per second. Video rate for persistent surveillance is not an option since data storage and data rates for downlink and bandwidth are all limited commodities in the military environment. Metrics that will evolve for PISR systems will likely be probability of target track over some distance on the ground or probability of target track for some given amount of time. Metrics for PISR will evolve into standards in the next few years.

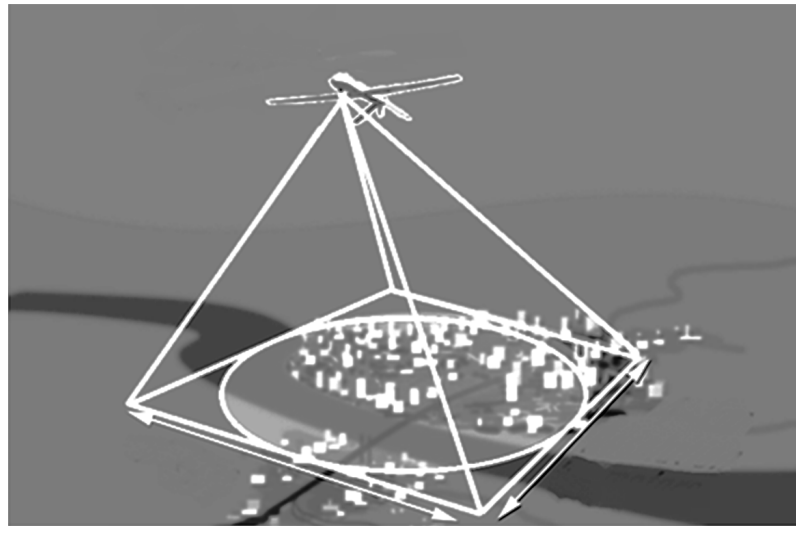


Figure 14.30 Persistent surveillance sensor coverage.

A few simple experiments [28] were performed at NVESD in 2008. Infrared simulations provided radiometrically (in the MWIR) correct human and vehicle data to a simulated PISR sensor. It was desired to understand the GSD and revisit rate needed to backtrack human targets. It was a simple experiment involving one moving human in a relatively simple background. The probability of target track as a function of GSD and revisit rate is shown in Figure 14.31.

The data showed that 1-m GSD and a 1-Hz revisit rate were sufficient for the simple task of a visual observer tracking a human. Under realistic conditions with many humans in the image along with complex backgrounds, the requirements can be between 0.25m and 0.5m with a revisit rate of 2 Hz. Reference [28] also

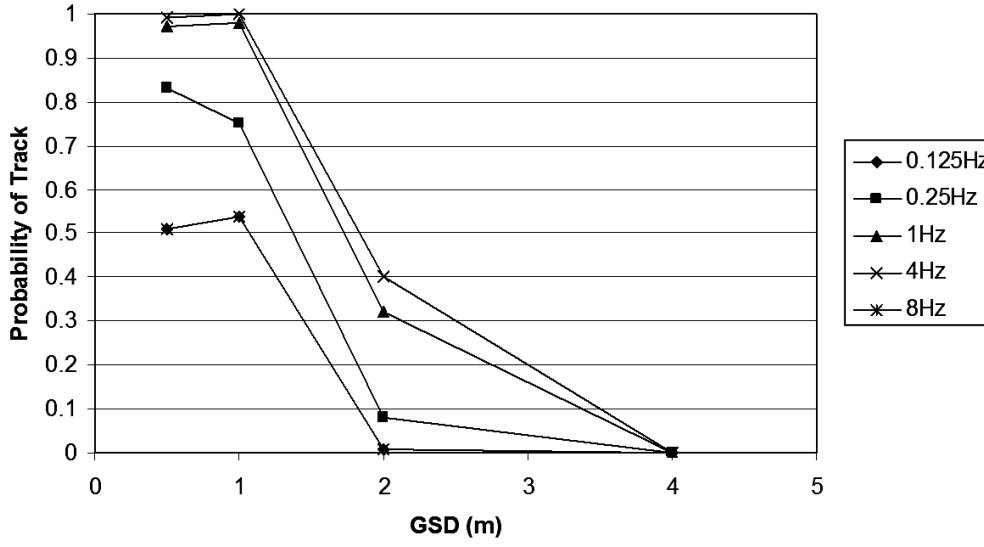


Figure 14.31 Simulated human track data as a function of GSD and revisit rate.

compares real imagery data to the simulated data where reasonable agreement is provided. Vehicle tracking is also simulated in Ref. [28] to provide some first estimates at the required GSD and revisit rate.

14.13 Exercises

14.1 Calculate the NETD for the imager in Example 14.1, given that the D* is 5 × 1011 Jones.

14.2 For the system described by the MTFs in Figure 14.9, describe which components limit the imager resolution. How would you increase the imager resolution?

14.3 The random spatial-temporal noise of a system is estimated at 0.037°K. Determine a reasonable estimate of the fixed row noise for a low-noise scanned system.

14.4 Plot the system CTF for the imager in Table 14.2 using a reasonable display assumption. Assume noise is not discernable (extremely) low noise.

References

[1] Waldman, G., and J. Wootton, *EO System Performance Modeling,* Norwood, MA: Artech House, 1993.

[2] Lloyd, J. M., "Fundamentals of Electro-Optical Imaging System Analysis" in the *Infrared and Electro-Optical Systems Handbook,* Bellingham, WA: ERIM and SPIE Press, 1993.

[3] Lloyd, J. M., *Thermal Imaging Systems,* New York City, NY: Plenum Press, 1975.

[4] Driggers, R., C. Halford, G. Boreman, and M. Wellfare, "Comparison of Two Frame Noise Calculations for Infrared Line Scanners," *Optical Engineering*, Vol. 29, 1990.

[5] Dereniak, E., and G. Boreman, *Infrared Detectors and Systems*, New York City, NY: Wiley, 1996.

[6] Shumaker, D., J. Wood, and C. Thacker, *FLIR Performance Handbook*, DCS Corporation, Washington DC, 1988.

[7] "Night Vision and Electronic Sensors Directorate Television Model," *U.S. Army CECOM Report*, Ft. Belvior, VA, 1994.

[8] Holst, G., *CCD Arrays, Cameras, and Displays*, Winter Park, FL: JCD Publishing, 1996.

[9] Scott, L., and J. D'Agostino, "Application of 3-D Noise to MRTD Prediction," Appendix C of FLIR92 Thermal Imaging Systems Performance Model, U.S. Army Night Vision and Electronic Sensors Directorate, Report 1993.

[10] Agostino, J., "A 3-D Noise Analysis Methodology," Appendix A of FLIR92 Thermal Imaging Systems Performance Model, U.S. Army Night Vision and Electronic Sensors Directorate, Report 1993.

[11] Webb, C., P. Bell, and G. Mayott, "Laboratory Procedure for the Characterization of 3-D Noise in Thermal Imaging Systems," *Proceedings of the IRIS Passive Sensors Symposium*, Laurel, MD, 1991.

[12] Vollmerhausen, R., Friedman, M., Reynolds, J., and Burks, S., "Modeling the blur associated with vibration and motion," *SPIE Proceedings*, Vol. 6543, 2007.

[13] Johnson, J., "Analysis of Image Forming Systems," Image Intensifier Symposium," Warfare Vision Branch, Warfare Electrical Engineering Dept., U.S. Army Engineering Research and Development Laboratories Report, Ft. Belvoir, VA, 1958.

[14] Ratches, J., "Static Performance Model for Thermal Imaging Systems," *Optical Engineering*, Vol. 15, No. 6, 1976, pp. 525–530.

[15] Moser, P., "Mathematical Model of FLIR Performance," NAVAIRDEVCEN Technical Memorandum NADC-20203: PMM, U.S. Naval Air Development Center, Warminster, PA, 1972.

[16] Vollmerhausen, R. H., E. Jacobs, and R. G. Driggers, "New metric for predicting target acquisition performance," *Opt. Eng.*, Vol. 43, No. 11, 2004, pp. 2806–2818.

[17] Preece, B. L., and J. P. Reynolds, "Improved noise model for the US Army Sensor Performance Metric," *SPIE Defense and Security Symposium*, Orlando, FL, 2011.

[18] Preece, B., J. Olson, J. Reynolds, and J. Fanning, "Performance Assessment of Treating Aliased Signal as Target Dependent Noise," *SPIE Defense and Security Symposium*, Orlando, FL, 2011.

[19] Barten, P. G. J., "Evaluation of the effect of noise on subjective image quality," *Proc. SPIE* 1452, pp. 2–15, 1991.

[20] Moyer, S., R. Teaney, J. Reynolds, J. Du Bosq, R. Thompson, et al., "Human target acquisition performance," *Military Sensing Symposium*, Pasadena, CA, 2012.

[21] Teaney, B., and J. Fanning, "Effect of image magnification on target acquisition performance," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing, Proc. SPIE*, 2008.

[22] Leachtenauer, J., W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General Image Quality Equation (GIQE)," *Applied Optics*, Vol. 36, No. 32, 1997.

[23] Leachtenauer, J. C., "National Imagery Interpretability Rating Scales: Overview and Product Description," *ASPRS/ASCM Annual Convention and Exhibition, Technical Papers*, Vol. 1, *Remote Sensing and Photogrammetry, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping*, Baltimore, MD, 1996, pp. 262–272.

[24] National Imagery and Mapping Agency, "General Image Quality Equation-Users Guide," Ver. 4.0, Data and Systems Analysis Division, Springfield, VA, 1996.

[25] Irvine, J., and J. Leachtenauer, "A Methodology for Developing Image Interpretability Ratings Scales," *ASPRS/ASCM Annual Convention and Exhibition, Technical Papers*, Vol. 1, *Remote Sensing and Photogrammetry, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping*, Baltimore, MD, 1996, pp. 273–281.

[26] Mohr, E., D. Hothem, J. Irvine, and C. Erdman, "The Multispectral Imagery Interpretation Rating Scale (MS IIRS)," *ASPRS/ASCM Annual Convention and Exhibition, Technical Papers*, Vol. 1, *Remote Sensing and Photogrammetry, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping*, Baltimore, MD, 1996, pp. 300–310.

[27] Leachtenauer, J, and R. Driggers, *Surveillance and reconnaissance imaging systems: modeling and performance prediction*, Norwood, MA: Artech House, 2001.

[28] Driggers, R., S. Aghera, P. Richardson, B. Miller, J. Doe, et al., "MWIR Persistent Surveillance Performance for Human and Vehicle Backtracking as a Function of Ground Sample Distance and Revisit Rate," *SPIE Defense and Security Symposium*, Orlando, FL, 2008.

CHAPTER 15

Search

In a military setting, infrared (IR) and EO systems are used for a variety of purposes. These include driver's aid, pilotage, aided target recognition, target identification, recognition, and detection. IR and EO systems allow vehicles and planes to get around covertly without using visible radiation. This defines the *driver's aid* and *pilotage* functions. *Aided target recognition* uses IR and EO system to put icons on a display that cue an observer to carefully examine part of the display for a target. The aim of aided target recognition is to more quickly determine where the targets are. *Target identification* is the ability to discriminate between specific targets. For example, an object in a display can be classified as a T72, T62, M1A, etc., tank. *Target recognition* is the ability to discriminate between objects by class. For example, an object in a display can be classified as being a truck, tank, or armored personal carrier. Target identification and recognition enable the classifying of an object as a threat or nonthreat and the prioritization of threats.

In this chapter the focus is on *dynamic detection*. We distinguish between three types of detection: *pure detection*, *discrimination detection*, and *dynamic detection*. *Pure detection* is a *static detection* process in which the target is perceived in a bland or uniform background. In pure detection the target location is not changing and no actual searching is performed; hence, use of the term *static*. *Discrimination detection*, also known as *Johnson detection*, is a search process in which the general location of the target is known. The difference between *discrimination detection* and *pure detection* is that in discrimination detection, the background is cluttered. Pure and discrimination detection are useful: They allow one to assess observer–sensor ability to acquire a target using several observers while avoiding statistical uncertainties caused by the fact that each observer searches in his/her individual way.

Dynamic detection or *search* requires acquiring or detecting the target when its location is not known (see Figure 15.1). Dynamic detection can be done in a single sensor FOV or in a field of regard (FOR) search that is typically much larger than the sensor FOV. The *field of regard* is the angular space to be searched. In dynamic detection, the sensor and the target can either be stationary or moving. In dynamic detection there can be one or multiple targets in the FOR and the atmospheric transmission can vary with position in the FOR and can also vary as a function of time. If the sensor has multiple fields of view, the observer is allowed to change to the field of view that is most appropriate. Dynamic detection includes the possibility of *cooperative detection*. In cooperative detection a number of observers are in communication with one another and the target is said to be detected when the first observer detects the target. In cooperative detection, the observers can be looking at individual displays with each display showing an image from a single sensor that is controlled by a single observer or each observer can control the sensor for their display. Search and material related to search is described in [1–53].

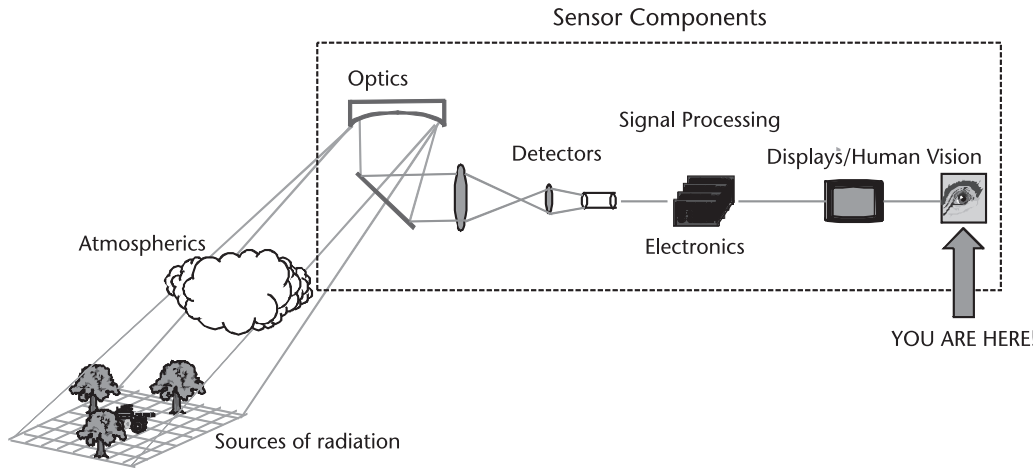


Figure 15.1 Human vision–brain component of the system.

15.1 Problem Definition

This section describes how a typical search experiment is done and exhibits the fundamental search equation. The problem we would like to address is several sensors on moving platforms searching a FOR for several military targets where some of the targets are stationary and some of the targets are moving. This is too hard a problem to start with and for that reason we consider a simpler problem: a single stationary target, say, a tank, is searched for by a single observer using a single sensor. Figure 15.2 shows a representative FOR and the search path of a particular soldier who searched for the single target within the scene.

One difficulty with a FOR search is that sometimes the target is not found because it was never in the FOV as the observer searched or was in the FOV for such a short time that the target was not found. As expected, a FOR search takes considerably longer than a FOV search. When evaluating sensor performance, in

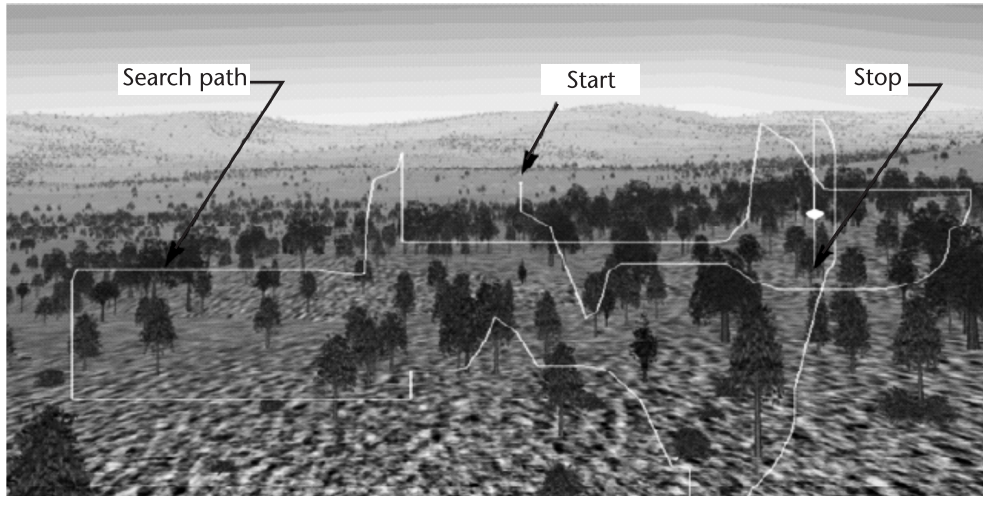


Figure 15.2 A representative field of regard image.

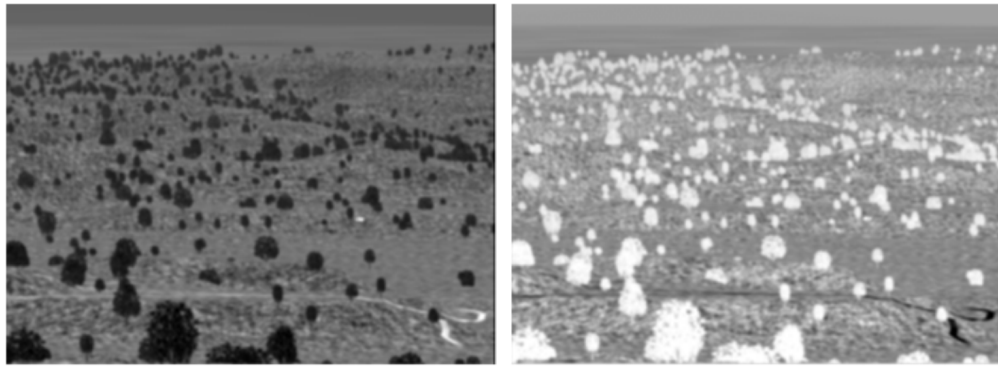


Figure 15.3 Field of view search with (left) low and (right) high clutter.

the interest of getting better statistics in less time, a FOV rather than FOR search is often done when only a few observers are available.

Figure 15.3 shows a FOV with low clutter and the same scene with high clutter. In the left part of the figure, the target is visible as a bright spot on an otherwise dark background. In the right part of the figure the target is still represented as a bright spot, but in this image it would be exceedingly difficult to identify it as a target.

15.2 Introduction to Search Theory

A group of observers is tasked to find a single target in an image on a monitor. Then

$$P(t) = P_\infty \left(1 - e^{-\frac{t}{\tau_{FOV}}} \right), \quad 0 \le t < \infty \tag{15.1a}$$

where

$P(t)$ = fraction of observers viewing the scene who find the target in a time less than or equal to t. $P(t)$ can also be thought of as target acquisition probability as a function of time

t = time the observers search the scene

P_∞ = fraction of observers who eventually find the target

τ_{FOV} = time constant that describes how quickly the target is found in FOV search.

Equation (15.1a) is shown graphically in Figure 15.4. Note that P_∞ is the asymptotic value of $P(t)$ and that τ_{FOV} describes how quickly the target is found. Equation (15.1a) reasonably asserts that the probability of finding the target is initially zero and monotonically increases to an asymptotic value.

Equation (15.1a) can be derived using an elementary model for the eye and probability theory. As an observer searches a FOV, the observer fixates points in the scene and moves from point to point in rapid movements called *saccades*. Fixation

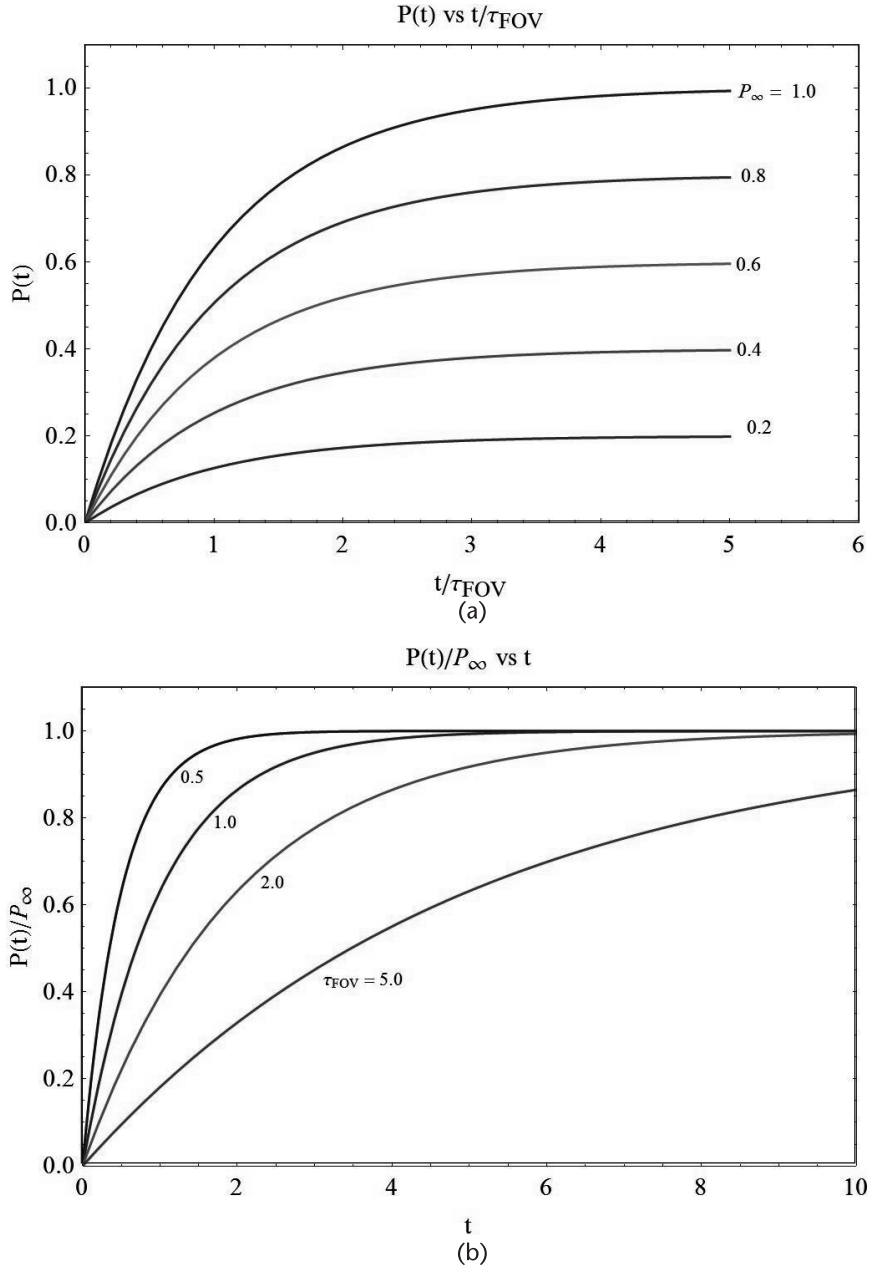


Figure 15.4 (a) Probability of target acquisition as function of t/τ_{FOV}. (b) As τ_{FOV} increases it takes longer for $P(t)$ to reach its asymptotic value.

times are variable and task dependent. A reasonable mean for a fixation time is 0.3 sec. In the derivation given here it is assumed that (1) fixation time is fixed at the mean value, (2) saccades are short compared to the fixation time and can be ignored, (3) the probability an observer finding the target in a fixation time is either zero or P_0, and (4) the observer searching the FOV is taken from an ensemble of identical observers taking part in the perception test.

To facilitate the derivation of (15.1a) the following symbols are introduced:

$P_{k,i}$ = probability the kth observer detects the target in the ith fixation
$\bar{P}_k(j)$ = probability the kth observer does not detect the target in j fixations
$P_k(j)$ = probability the kth observer does detect the target in j fixations
$P_k(t)$ = probability the kth observer detects the target in time t
P_k = probability the kth observer detects the target in a fixation; that is, assume $P_{k,i}$ is independent of i
$\tau_{FOV,k}$ = time constant for the kth observer
τ_{FOV} = time constant for the ensemble of observers
t_f = mean fixation time
n = number of observers in the ensemble
m = number of observers who detect the target.

Assuming that the probabilities for not finding the target are independent, the probability $\bar{P}_k(j)$ that the kth observer *does not* detect the target in j fixations is

$$\bar{P}_k(j) = \prod_{i=1}^{j} (1 - P_{k,i}) \tag{15.2a}$$

which implies that the probability the kth observer *does* detect the target in j fixations is

$$P_k(j) = 1 - \prod_{i=1}^{j} (1 - P_{k,i}) \tag{15.2b}$$

If the $P_{k,i}$ is not a function of i, then write $P_{k,i} = P_k$ and (15.2b) becomes

$$P_k(j) = 1 - (1 - P_k)^j = 1 - e^{\ln(1-P_k)^j} = 1 - e^{j\ln(1-P_k)} \tag{15.2c}$$

The elapsed time t, the fixation time t_f, and j are related:

$$j = \frac{t}{t_f} \tag{15.2d}$$

Using this result, (15.2c) becomes

$$P_k(t) = 1 - e^{\frac{t}{t_f} ln(1-P_k)} \tag{15.2e}$$

Looking up the Taylor series for $ln(1 - P_k)$, we find

$$ln(1 - P_k) = -\left(P_k + \frac{P_k^2}{2} + \frac{P_k^3}{3} + \ldots \right) \tag{15.2f}$$

Since P_k is the probability of detecting the target in a single fixation, it is expected that $P_k << 1$ and to an excellent approximation:

$$ln(1 - P_k) = -P_k \tag{15.2g}$$

Using this result, (15.2e) becomes

$$P_k(t) = 1 - e^{-\frac{P_k}{t_f} t} = 1 - e^{-\frac{t}{\tau_{FOV,k}}} \quad \text{where} \quad \tau_{FOV,k} \equiv \frac{t_f}{P_k} \tag{15.2h}$$

This equation shows that $\tau_{FOV,k}$ can be interpreted as the ratio of the mean fixation time t_f to the probability P_k. The derivation of (15.2h) is based on the performance of the kth observer.

Now focus on the implications of (15.2h) for an ensemble of observers. In (15.2h) P_k denotes the probability of detecting the target in any fixation for the kth observer. Some members of the ensemble never acquire the target and some members do acquire the target. For those observers who do not acquire the target, P_k in (15.2h) is evidently equal to zero. For simplicity it is assumed that all observers who do acquire the target have parameter P_k in (15.2h) equal to P_0. The value of $P(t)$ is obtained by summing (15.2h) over all observers and dividing by the number of observers:

$$P(t) = \frac{1}{n}\sum_{k=1}^{n} P_k(t) = \frac{1}{n}\sum_{k=1}^{n}\left(1 - e^{-\frac{P_k}{t_f}t}\right) = \frac{m}{n}\left(1 - e^{-\frac{P_0}{t_f}t}\right), \quad 0 \le t < \infty \tag{15.2i}$$

Comparing this result with (15.1a), $P\infty$ and can be expressed in terms of m, n, P_0, and t_f:

$$P_\infty = m/n \text{ and } \tau_{FOV} = t_f/P_0 \tag{15.2j}$$

This concludes the derivation of (15.1a).

A cumulative distribution function (CDF) and probability density function (pdf) are implied by (15.1a). The CDF associated with (15.1a) is

$$P_c(t) = 1 - e^{-t/\tau_{FOV}}, \quad 0 \le t < \infty \tag{15.3a}$$

Note that $P_c(t)$ has the requisite property of a CDF: $P_c(0) = 0$ and $P_c(\infty) = 1$. The observation that $P_c(\infty) = 1$ implies that the CDF and pdf apply only to the observers who find the target.

Recall from probability theory that [28, 41, 42] the pdf is obtained by differentiating the CDF. We use the convention that probability density functions are denoted by lowercase letters and cumulative distribution functions are denoted by uppercase letters.

Differentiating (15.3a) with respect to time yields $p(t)$, the pdf associated with (15.1a) and (15.3a):

$$p(t) = \frac{1}{\tau_{FOV}} e^{-t/\tau_{FOV}}, \quad 0 \le t < \infty \tag{15.4a}$$

Equation (15.4a) allows us to easily estimate τ_{FOV} from experimental measurements. Let <t> denote the mean time to detect the target. Then,

$$< t > = \int_0^\infty t\, p(t)\, dt = \frac{1}{\tau_{FOV}} \int_0^\infty t e^{-t/\tau_{FOV}}\, dt = \tau_{FOV} \tag{15.5a}$$

Equation (15.5a) has a discrete representation. If t_i denotes the time the ith observer detects a target and m observers detect the target, then

$$\tau_{FOV} = \frac{1}{m}\sum_{i=1}^{m} t_i \tag{15.5b}$$

The development up to this point has been based on the validity of (15.1a). Perception tests show that (15.1a) needs to be modified to include a delay time t_d:

$$P(t) = \begin{cases} 0, & t < t_d \\ P_\infty \left(1 - e^{-\frac{t-t_d}{\tau_{FOV}}}\right), & t_d \le t < \infty \end{cases} \tag{15.1b}$$

The delay time can be thought of as the sum of a *perception time*, a *decision time*, and a *reaction time*. When an observer looks at a scene it takes a finite amount of time, called *perception time*, to understand the scene. In FOV search, to make the detection task challenging, the target frequently has few *V* cycles on it. As a result, in a forced choice experiment, an observer cannot immediately determine that the object of interest is a target, and it is often necessary to carefully examine other points of interest before declaring a target. Time spent deciding if a point of interest is a target is called the *decision time*. After an observer decides an object of interest is a target, it takes time to move a mouse-activated pointer to the target and then click to record the target location and the time the target is detected. This is called the *reaction time*.

There is no established theory of the delay time and it is possible other factors contribute to the delay time. Although t_d may depend on P_∞ and τ_{FOV}, typically it is described as a constant, probably because the errors that occur when measuring t_d mask any dependence t_d may have on P_∞ and τ_{FOV}.

If we accept the validity of (15.1b), then (15.3a) and (15.4a) need to be changed:

$$P_c(t) = \begin{cases} 0\,, & t < t_d \\ 1 - e^{-\frac{t-t_d}{\tau_{FOV}}}\,, & t_d \le t < \infty \end{cases} \tag{15.3b}$$

$$p(t) = \begin{cases} 0\,, & t < t_d \\ \dfrac{1}{\tau_{FOV}} e^{-\frac{t-t_d}{\tau_{FOV}}} & t_d \le t < \infty \end{cases} \tag{15.4b}$$

Equation (15.5a) also needs to be changed to accommodate (15.1b):

$$< t > = \tau_{FOV} + t_d \tag{15.5c}$$

If the search data need to be described by (15.1b), then typically it is more practical to determine τ_{FOV} from a least-squares fit to the data rather than use (15.5b).

15.3 Technique for Estimating Search Parameters and Their Uncertainties

In this section we show how to estimate the search parameters P_∞ and τ_{FOV} from perception experiment data.

Example 15.1

Twenty observers are given as much time as they want to search a FOV that contains a single target. Of these 20 observers, 17 find the target. Given the detection

times (units are seconds) of the observers listed below, (a) estimate P_∞ and τ_{FOV} and (b) construct a graph showing how the probability for target acquisition grows with time and compare it with the model result.

```
{0.10811, 0.905907, 0.980514, 1.0507, 1.42483,
 1.4463, 2.23957, 3.08201, 3.36015, 3.57875, 3.68921,
 4.52383, 5.20406, 5.97654, 7.90869, 9.26624, 14.4032}
```

Solution: (a) Equation (15.2j) implies that for this field of view $P_\infty = 17/20 = 0.85$. Equation (15.5b) implies that τ_{FOV} is estimated by calculating the mean detection time. Doing this we find $\tau_{FOV} = 4.06$ sec. (b) Because there are 20 observers, each time an observer detects a target the probability for target detection increases by 1/20 = 0.05. Pairing detection times with probability of target acquisition we get the following table:

```
{{0.10811, 0.05}, {0.905907, 0.1}, {0.980514, 0.15},
 {1.0507, 0.2}, {1.42483, 0.25}, {1.4463, 0.3}, {2.23957, 0.35},
 {3.08201, 0.4}, {3.36015, 0.45}, {3.57875, 0.5},
 {3.68921, 0.55}, {4.52383, 0.6}, {5.20406, 0.65},
 {5.97654, 0.7}, {7.90869, 0.75}, {9.26624, 0.8}, {14.4032, 0.85}}
```

The experimental data, denoted by dots in Figure 15.5, are consistent with (15.1a) and this supports the use of (15.5b) to calculate τ_{FOV}. The smooth curve compares (15.1) calculated with $P_\infty = 0.85$ and $\tau_{FOV} = 4.06$ with the experimental data.

Meitzler[34] states "Tank commanders or fighter pilots, for example, can often relate their relative success in tactical engagements to the time interval necessary

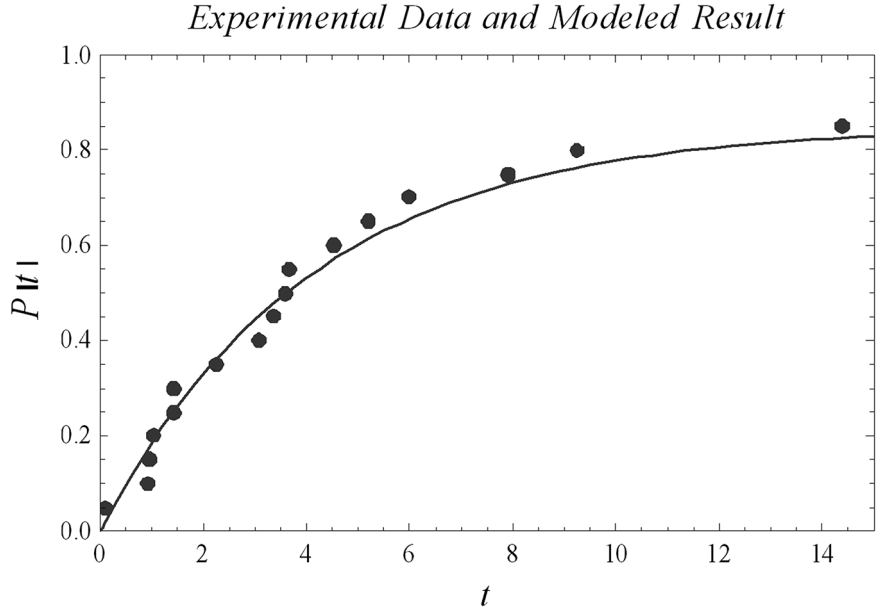


Figure 15.5 Experimental data and modeled result for Example 15.1.

to acquire the target." Obviously the probability of acquiring the target is also important to the outcome of a tactical engagement. For this reason, it is important to measure, τ_{FOV}, P_∞, and the uncertainties in these measurements.

We now show that the standard deviation for the measurement of P_∞ is

$$\sigma_{P_\infty} = \sqrt{\frac{P_\infty(1-P_\infty)}{n}} \tag{15.6}$$

where n is the number of observers who take part in the perception experiment and P_∞ is estimated using (15.2j). Let m denote the number of observers who detect the target when there are n observers. Then (15.2j) implies

$$P_\infty = \frac{m}{n} \quad \Rightarrow \quad \sigma_{P_\infty} = \frac{\sigma_m}{n} \tag{15.7a}$$

Let $p(m)$ denote the probability of having m detections in n draws where the probability of detection on any draw is P_∞. Then $p(m)$ is described by the binomial probability distribution function and is given by

$$p(m) = C_m^n P_\infty^m (1-P_\infty)^{n-m} \quad \text{where } C_m^n \equiv \frac{n!}{m!(n-m)!} \tag{15.7b}$$

The standard deviation for the number of successes in a binomial distribution is well known[41]:

$$\sigma_m = \sqrt{nP_\infty(1-P_\infty)} \tag{15.7c}$$

Equations (15.7a) and (15.7c) imply the result of (15.6).

Sometimes the fractional error is the pertinent quantity. Divide both sides of (15.6) by P_∞:

$$\frac{\sigma_{P_\infty}}{P_\infty} = \sqrt{\frac{(1-P_\infty)}{nP_\infty}} \tag{15.8}$$

Equations (15.6) and (15.8) are the modeled result. In deriving these equations P_∞ is the true value for this parameter. However, in using these equations we can approximate the true value of P_∞ by the measured value.

Observe in (15.6) that when $P_\infty = 0$ or 1, $\sigma_{P_\infty} = 0$. This is just what is expected. When $P_\infty = 0$, no observer finds the target and the standard deviation is zero. Similarly, when $P_\infty = 1$, all observers find the target and the standard deviation is zero. Note from (15.8) that the fractional error goes down with the square root of the number of observers as expected. A graph of (15.6) is shown in Figure 15.6.

We will now show that the standard deviation for τ_{FOV} is

$$\sigma_{\tau_{FOV}} = \frac{\tau_{FOV}}{\sqrt{m}} = \frac{\tau_{FOV}}{\sqrt{P_\infty n}} \tag{15.9}$$

where m and n are, respectively, the number of observers who detect the target and the number of observers who take part in the perception experiment. In (15.9), the second equality follows from the first using (15.2j).

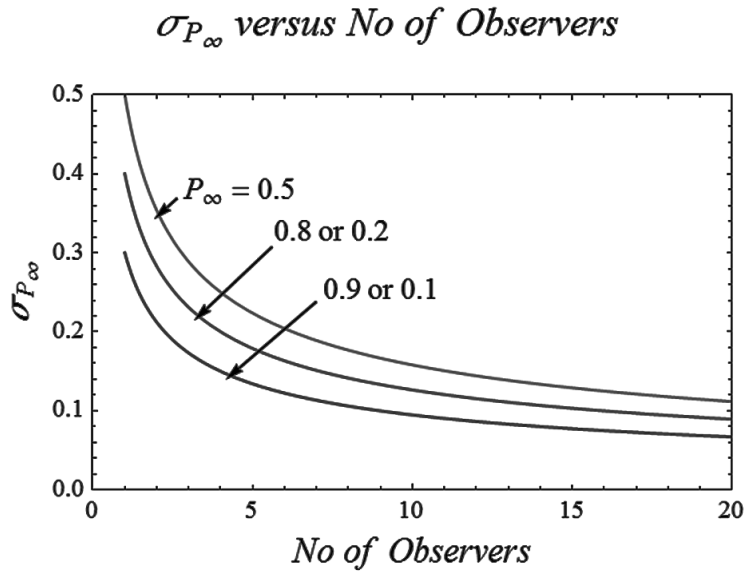


Figure 15.6 Standard deviation in P_∞ versus number of observers and P_∞. The relationship is depicted by solid lines even though the number of observers is an integer value.

To derive (15.9), imagine that we have a large number of identical observers and that the same scene is repeatedly shown to groups of n observers. Using (15.5b), on the ith trial we get

$$\tau_{FOV_i} = \frac{1}{m_i}(t_1 + t_2 + \cdots + t_{m_i}) \tag{15.10a}$$

where the detection times t_k are random variables taken from an exponential pdf of width τ_{FOV}. Of course, m_i the number of observers who detect the target in the ith trial is also a random variable. The τ_{FOV} of (15.10a) are averaged to estimate τ_{FOV} and the problem is to compute the standard deviation of τ_{FOVi}. Observe that m_i and $t_1 + t_2 + \ldots$ in (15.10a) are both random variables that contribute to the variation in τ_{FOV}. Here, the variation in τ_{FOV} due to the variation in m_i is neglected. Thus, m_i is fixed at its expected value $m = E[m_i]$. Then

$$\tau_{FOV_i} = \frac{1}{m}(t_1 + t_2 + \cdots + t_m) \tag{15.10b}$$

Take the log of this equation:

$$ln(\tau_{FOV_i}) = ln(t_1 + t_2 + \cdots + t_m) - ln(m) \tag{15.10c}$$

In each trial P_∞, τ_{FOV}, and m are constant. To find the relationship between $\delta\tau_{FOV_i}$ and $\delta(t_1 + t_2 + \ldots + t_m)$, differentiate (15.10c) keeping m constant:

$$\frac{\delta\tau_{FOV_i}}{\tau_{FOV_i}} = \frac{\delta(t_1 + t_2 + \cdots + t_m)}{t_1 + t_2 + \cdots + t_m} \tag{15.10d}$$

This equation describes how an error in $t_1 + t_2 + \ldots + t_m$ produces an error in τ_{FOV_i}. It also provides guidance on how $\sigma_{\tau_{FOV}}$ and $\sigma_{t_1+t_2+\ldots+t_m}$ are related:

$$\frac{\sigma_{\tau_{FOV}}}{\tau_{FOV}} = \frac{\sigma_{t_1+t_2+\cdots+t_m}}{E[t_1+t_2+\cdots+t_m]} \tag{15.10e}$$

The numerator of (15.10e) is readily computed:

$$\sigma^2_{t_1+t_2+\ldots+t_m} = \sigma^2_{t_1} + \ldots + \sigma^2_{t_m} = m\,\sigma^2_{\tau_{FOV}} = m\,\tau^2_{FOV} \tag{15.10f}$$

In (15.10f) the first equality follows since the detection times t_k are not correlated. The second equality is true since each t_k is drawn from an identical distribution of width $\sigma_{\tau_{FOV}}$. The third equality follows since the variance of an exponential pdf is τ^2_{FOV}. The denominator of (15.10e) is readily computed:

$$E[t_1+t_2+\cdots+t_m] = E[t_1]+\cdots+E[t_m] = m\,\tau_{FOV} \tag{15.10g}$$

where the first equality follows from the linearity of the expectation operator and the second equality follows since the expectation of each t_k is τ_{FOV}. Using (15.10f) and (15.10g), (15.10e) becomes

$$\frac{\sigma_{\tau_{FOV}}}{\tau_{FOV}} = \frac{\sqrt{m}\tau_{FOV}}{m\tau_{FOV}} = \frac{1}{\sqrt{m}} \tag{15.10h}$$

Solving (15.10h) for $\sigma_{\tau_{FOV}}$ and realizing that $m = P_\infty n$, we get (15.9), a graph of which is shown in Figure 15.7.

Equation (15.9) is reasonable. The quantity $\sigma_{\tau_{FOV}}$ has the dimensions of time and the only things $\sigma_{\tau_{FOV}}$ can depend on are τ_{FOV}, P_∞, m, and n. The only way $\sigma_{\tau_{FOV}}$ can depend on τ_{FOV} is to be proportional to it. One expects $\sigma_{\tau_{FOV}}$ to go down as the square root of the number of events used to estimate τ_{FOV} and this implies $\sigma_{\tau_{FOV}}$ goes down as the square root of $P_\infty n$.

15.4 Search Parameters and NV-IPM

In Section 15.3 we showed how the search parameters P_∞ and τ_{FOV} can be determined in FOV search experiments. In this section we indicate how the search parameters can be estimated from the NV-IPM model discussed in Chapters 13 and 14. Equation (13.42b) gives the probability of detecting the target as a function of $V(R)$, which depends on the sensor–target range R and an empirically determined parameter V_{50}. *The quantity P(R) in (13.42b), calculated using the equations of Chapters 13 and 14, is the P_∞ in (15.1a) and (15.1b).*

We expect that when the targets are easier to find, P_∞ will increase and the targets will be found in less time. As P_∞ increases, it is expected that τ_{FOV} will decrease and this is observed in perception tests done in a lab using imagery collected in the field. A variety of results that show the relationship between τ_{FOV} and P_∞ have been given. Howe[27] gives the following result:

$$\tau_{FOV} = \frac{3.4}{P_\infty}, \quad P_\infty \le 0.90 \tag{15.11a}$$

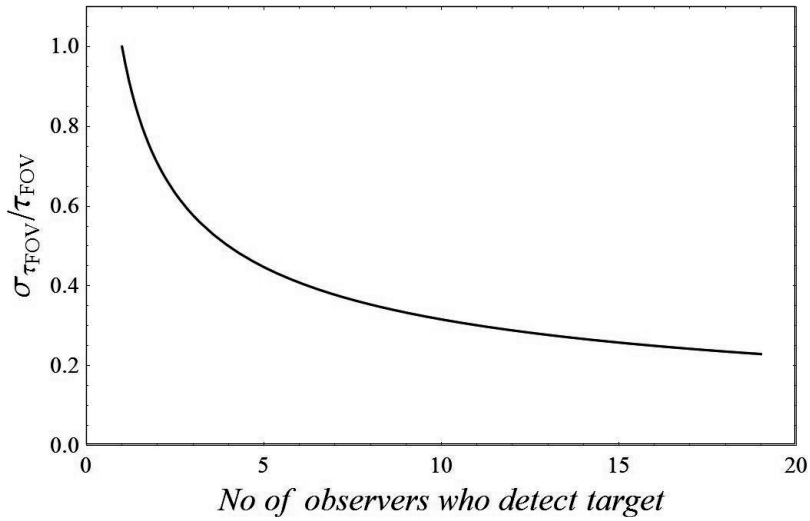


Figure 15.7 Graphical illustration of (15.9). The relationship is depicted by a solid line even though the number of observers who detect the target is an integer.

where it is asserted that this equation is valid for low to medium clutter situations, given unlimited search time. Note that this relationship between τ_{FOV} and P_∞ and similar relationships given below hold for the ensemble of targets and observers used in the experiment and cannot be expected to accurately predict the result for any particular target or observer. Edwards[16] give these results:

$$\tau_{FOV} = 3.0 - 2.2P_\infty \tag{15.11b}$$

$$\tau_{FOV} = 4.0 - 2.65P_\infty \tag{15.11c}$$

where (15.11b) refers to an experiment done with computer-generated paint-the-night imagery in moderate clutter; (15.11c) corresponds to experiments done with field imagery. In the experiments used to determine (15.11b) and (15.11c), the observers were constrained in how much time they were allowed to search the FOV. No such constraints were in place in the experiments used to determine (15.11a). In subsequent time-limited search experiments, Edwards and colleagues[17] refined their model and reported

$$\tau_{FOV} = 3.5 - 2.5P_\infty \tag{15.11d}$$

The imagery used to derive (15.11a) through (15.11d) involved perception experiments using real or simulated IR imagery in the longwave band that involved military vehicles in a rural environment. Devitt[5] reported time-limited search results for human targets using MWIR and LWIR sensors in an urban environment:

$$\text{MWIR day: } \tau_{FOV} = 3.82 - 2.51P_\infty \tag{15.11e}$$

$$\text{LWIR day: } \tau_{FOV} = 3.13 - 2.09P_\infty \tag{15.11f}$$

$$\text{MWIR night: } \tau_{FOV} = 4.82 - 4.13P_\infty \tag{15.11g}$$

$$\text{LWIR night: } \tau_{FOV} = 2.40 - 1.174P_\infty \tag{15.11h}$$

Equation 15.11a uses (15.1a) as the search model. Equations (15.11b) through (15.11h) use (15.1b) as the search model. All of these equations were obtained by determining P_∞ and τ_{FOV} in perception experiments; P_∞ and τ_{FOV} pairs are graphed and the relationship between them is found using a fitting procedure. Graphs of (15.11b) through (15.11h) are shown in Figure 15.8. Equation (15.11a) was not plotted in Figure 15.8 because for this case τ_{FOV} is considerably larger than the other estimates and does not conveniently fit on the graph.

Although Figure 15.8 indicates that τ_{FOV} is confined to a fairly narrow range of values for any value of P_∞—and this may be the case for scenes encountered in a military context—this is not the case in general as indicated by the *Waldo* books.[23, 24] In the *Waldo* books, Waldo and each character are very clearly defined but in any FOV there are hundreds of characters who are similar to although clearly different from Waldo. The determined searcher will nearly always find Waldo with the result that P_∞ is near one, but the time required to find Waldo is typically measured in units of minutes rather than seconds.

Equations (15.11b) through (15.11h) all have the form

$$\tau_{FOV} = \alpha - \beta P_\infty \tag{15.11i}$$

For a FOV without a target, $P_\infty = 0$. Then using (15.5c) we conclude that

$$< t >_{EFOV} - t_d = \alpha \quad \Rightarrow \quad < t >_{EFOV} = \alpha + t_d \tag{15.12}$$

where $< t >_{EFOV}$ is the mean time spent searching an empty field of view. As we will see subsequently, the mean time spent searching an empty FOV is important for describing FOR search.

We hypothesize that α depends on (1) the clutter in the scene, (2) the degree to which the observer believes a target is present, (3) observer fatigue, and (4) observer patience. We hypothesize that β depends on (1) target size and contrast and (2) clutter. Experiments to understand how α and β depend on the scene have not been done.

The important points of this section can be summarized as follows:

1. P_∞ can be estimated using the methods of Chapters 13 and 14.
2. Estimates for P_∞ made using these equations have been tested mainly with FOV search but are thought to also apply to FOR search.
3. The value for τ_{FOV} can be estimated from P_∞ but the particular relationship depends on clutter, the sensor, and the required task.
4. Note that in (15.11a) through (15.11i) τ_{FOV} and P_∞ apply to the ensemble of targets used in the perception experiment and cannot be applied with confidence to any particular target.
5. There is no theory relating τ_{FOV} to P_∞, so currently the relationship between these quantities must be determined empirically for a particular scenario.

15.5 Time-Limited Search

Many of the experiments relating τ_{FOV} to P_∞ described in the preceding section refer to search with unlimited time. In this section we define *time-limited search* and the equations that describe it.

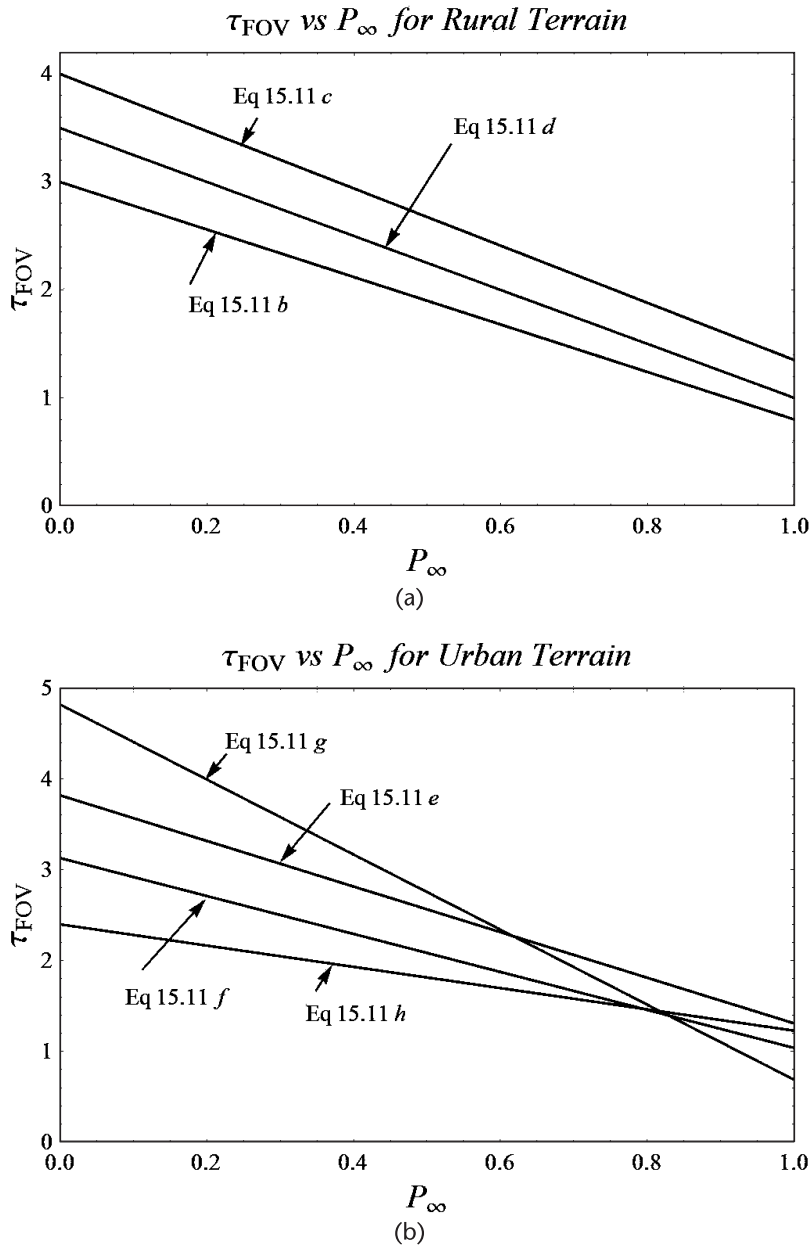


Figure 15.8 (a) Empirical relationships between the ensemble average τ_{FOV} and P_∞ for military vehicles in rural terrain. (b) Empirical relationships between the ensemble average τ_{FOV} and P_∞ for human targets in urban terrain.

In time-limited search the observers are given a limited period of time to find a target. Representative[16,17] times the observer is allowed to search for a target are 3, 6, 9, 12, and 17 sec. For example, in FOV search, if the observer is given 6 sec then for 6 sec the FOV image is displayed on a monitor; after 6 sec the monitor becomes blank and the observer no longer has an opportunity to find the target. Let t_L denote the amount of time the observer is allowed to look at an image. Then for $t > t_L$ the pdf must be zero and the generalization of (15.4a) is

$$p(t) = \begin{cases} \dfrac{C}{\tau_{FOV}} e^{-\frac{t}{\tau_{FOV}}} & 0 \le t < t_L \\ 0 & t_L < t \end{cases} \tag{15.13a}$$

The constant C is determined by the requirement that the integral of the pdf over all allowed values of t is one:

$$\int_0^{t_L} p(t)\, dt = 1 \tag{15.14}$$

Then (15.13a) becomes

$$p(t) = \begin{cases} \dfrac{1}{\tau_{FOV}} \dfrac{1}{1 - \exp\left(-\dfrac{t_L}{\tau_{FOV}}\right)} e^{-\frac{t}{\tau_{FOV}}} & 0 \le t < t_L \\ 0 & t_L < t \end{cases} \tag{15.13b}$$

Figure 15.9 compares the time-limited pdf of (15.13b) with the time-unlimited pdf of (15.4a). Note the similarity in the shape of the time-limited and time-unlimited pdfs. Over the region where the time-limited pdf is defined, it is greater than that of the time-unlimited pdf so that the integral of both density functions is one.

The CDF is the integral of the pdf [28, 41, 42]:

$$P_c(t) = \int_0^t p(t)\, dt \tag{15.15}$$

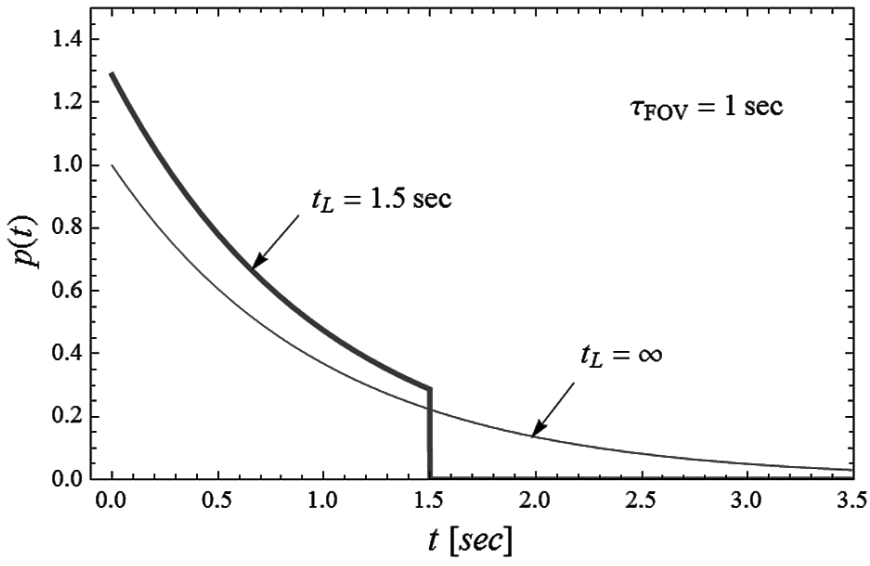


Figure 15.9 Probability density functions for time-limited and time-unlimited search.

Substituting (15.13b) into (15.15) we get the generalization of (15.3a):

$$P_c(t) = \begin{cases} \dfrac{1 - e^{-t/\tau_{FOV}}}{1 - e^{-t_L/\tau_{FOV}}} & 0 \le t \le t_L \\ 1 & t_L < t \end{cases} \tag{15.16}$$

A comparison of the CDF curves for time-limited and time-unlimited search is shown in Figure 15.10. Note from (15.16) or from Figure 15.10 that when $t = t_L$, the CDF is one.

In time-limited search, P_∞ is still obtained from perception experiments using (15.2j). The generalization of (15.1a) for time-limited search is obtained by multiplying (15.15) by the fraction of observers who find the target in time t_L, denoted here as P_{t_L}:

$$P(t) = \begin{cases} P_{t_L} \dfrac{1 - e^{-t/\tau_{FOV}}}{1 - e^{-t_L/\tau_{FOV}}} & 0 \le t \le t_L \\ P_{t_L} & t_L < t \end{cases} \tag{15.17}$$

The mean time for detection $< t >$ is obtained by integrating the pdf as in (15.5a) only now the pdf is given by (15.13b). The result is

$$< t > = \tau_{FOV} - \frac{t_L}{e^{t_L/\tau_{FOV}} - 1} \tag{15.18}$$

All of the time-limited equations given so far are for the case where t_d is zero or can be neglected. If t_d cannot be neglected, then the generalization of (15.13b) is

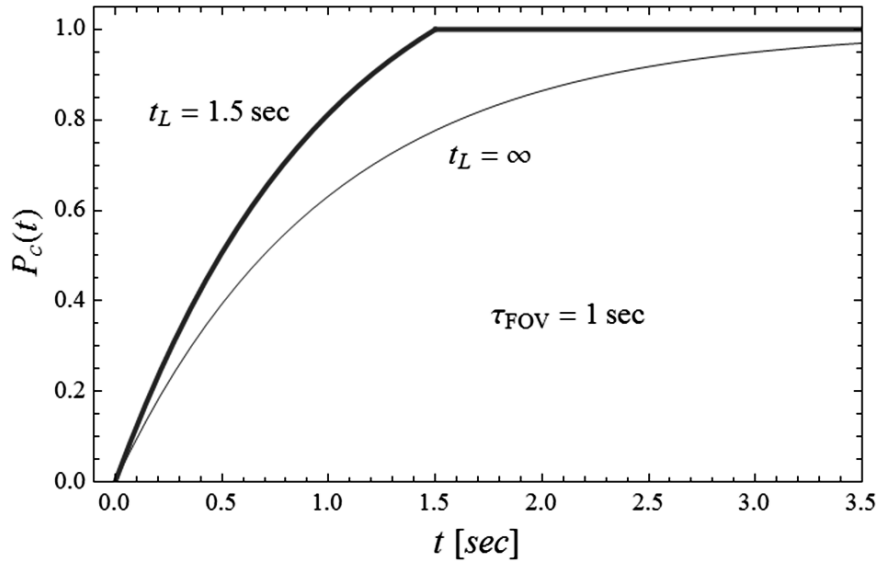


Figure 15.10 Cumulative distribution functions for time-limited search and time-unlimited search.

$$p(t) = \begin{cases} \dfrac{1}{\tau_{FOV}} \dfrac{1}{1 - \exp\left(-\dfrac{t_L - t_d}{\tau_{FOV}}\right)} e^{-\frac{t - t_d}{\tau_{FOV}}} & t_d \le t < t_L \\ 0 & t_L < t \end{cases} \tag{15.19}$$

This equation is illustrated in Figure 15.11 for the case where t_d is either 0.0 or 0.3 sec with τ_{FOV} equal to 1.0 sec and t_L equal to 1.5 sec. Note that the effect of a nonzero t_d is to shorten the interval over which the pdf is defined but to otherwise leave the shape unchanged. For $0.3 < t < 1.5$, the pdf with the nonzero value for t_d is greater than the one with a zero value for t_d and that must be true since the integral over both pdfs must equal one.

If t_d cannot be neglected, the generalization of (15.16) is

$$P_c(t) = \begin{cases} 0 & t < t_d \\ \dfrac{1 - e^{-(t - t_d)/\tau_{FOV}}}{1 - e^{-(t_L - t_d)/\tau_{FOV}}} & t_d \le t \le t_L \\ 1 & t_L < t \end{cases} \tag{15.20}$$

This equation is illustrated in Figure 15.12.

If t_d cannot be neglected, the generalization of (15.17) is

$$P(t) = \begin{cases} 0 & t < t_d \\ P_{t_L} \dfrac{1 - e^{-\frac{t - t_d}{\tau_{FOV}}}}{1 - e^{-\frac{t_L - t_d}{\tau_{FOV}}}} & t_d \le t \le t_L \\ P_{t_L} & t_L < t \end{cases} \tag{15.21}$$

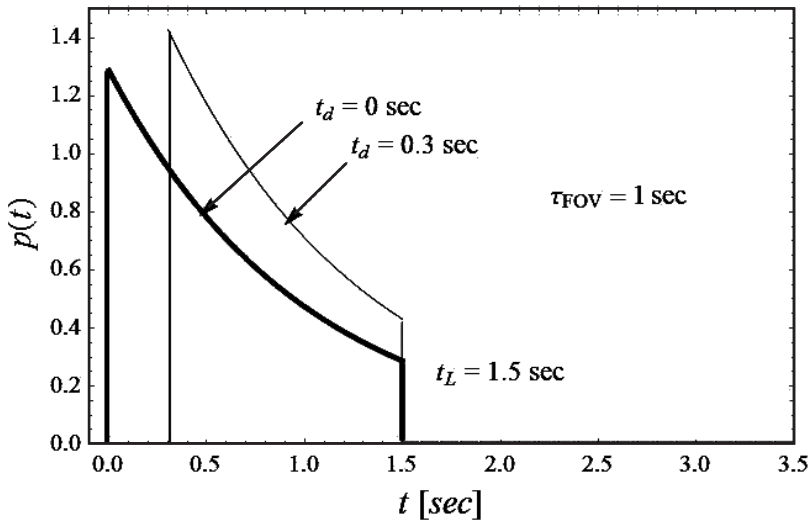


Figure 15.11 Comparison of time-limited pdfs with delay times of 0.0 and 0.3 sec.

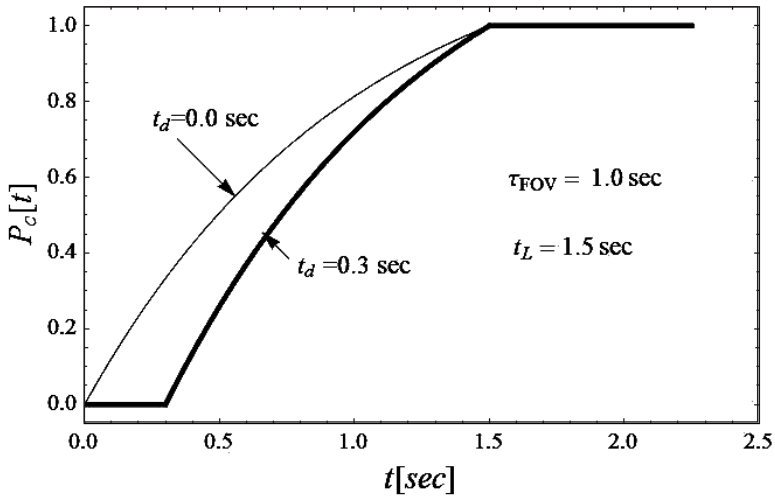


Figure 15.12 Comparison of time-limited CDFs with delay times of 0.0 and 0.3 sec.

For the case where t_d cannot be neglected, the generalization of (15.18) is

$$< t > = \tau_{FOV} + t_d - \frac{t_L - t_d}{e^{(t_L - t_d)/\tau_{FOV}} - 1} \tag{15.22}$$

15.6 Field of Regard Search

In this section we generalize[3] the FOV results found in Section 15.2 to FOR searches.

With unlimited time, the usual assumption is that P_∞ in FOR search is the same as that in FOV search and is the $P(R)$ calculated using (13.42b). As the FOR increases, but includes the FOV that contains the target, it may be that P_∞ decreases. The reason for this is that in a sufficiently large FOV some observers may give up looking for the target and some of those observers may have found the target in a smaller FOV.

Other important questions concerning FOR searches include "How does the probability of target acquisition grow with time?" and "How does the mean time to find the target depend on the size of the FOR?" One answer for how the probability of target acquisition grows with time is

$$P(t) = P_\infty(1 - e^{-t/\tau_{FOR}}), \quad 0 \le t < \infty \tag{15.23}$$

One estimate for τ_{FOR} is

$$\tau_{FOR} = n\,\tau_{FOV} \tag{15.24}$$

where n is the number of fields of view needed to cover the FOR. The rationale for (15.24) is this: Typically the target is hard to find in any particular FOV, so it is reasonable to think that the amount of time spent searching an empty FOV is comparable to the amount of time spent searching a FOV that contains the target.

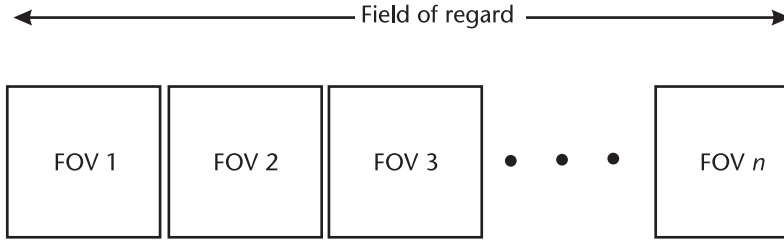


Figure 15.13 Model for describing FOR searches.

Comparing (15.24) with (15.5a), we see that (assuming t_d is negligible) the mean time to detect a target is proportional to the product of n, which describes how large the FOR is, and τ_{FOV} which describes how long it takes to find a target in a FOV. For hard-to-find targets, (15.24) is a reasonable result but it is not a reasonable result for an easy-to-find target.

To improve on the results of (15.23) and (15.24), we consider the model depicted in Figure 15.13. Assume the field of regard is covered by n fields of view and that the observer spends time t_0 examining each FOV. Conceptually this could be done by automating the sensor and having it stare at each FOV for a time t_0 and then instantaneously move to the next FOV. When the sensor gets to the last FOV, it instantaneously returns to the first FOV. We also assume unlimited search time and that a single target is placed randomly in one of the fields of view so the probability of the target being in any one FOV is $1/n$. The time t_0 is arbitrary, is chosen in advance, and is input into the mount that step-stares the sensor over the FOR. Although conceptually, t_0 can have any positive value, a reasonable value for this parameter is about 2 sec.

Typically, searching is done by continuously moving the sensor over the FOR in a random fashion similar to that shown in Figure 15.2. Here it is assumed that the mean time to find the target with random search will be comparable to the search process depicted in Figure 15.13. The value in using the search process described in Figure 15.13 is that it is easier to model than the search method of Figure 15.2.

The model described in Figure 15.13 yields a mean time τ_{FOR} to find the target that is given by the relationship

$$\tau_{FOR} = \frac{(2-P_s)}{2P_s}(n-1)t_0 + \tau_{FOV};\ \ P_s \equiv 1 - e^{-t_0/\tau_{FOV}} \tag{15.25}$$

The derivation of this formula is intricate and is deferred to Appendix 15A at the end of this chapter.

Before examining the derivation of this formula, let's see if it has the correct limiting behavior. Suppose n =1, which corresponds to the case where the size of the FOR and FOV is the same. In that case (15.25) yields the result that τ_{FOR} equals τ_{FOV} as expected. Suppose $\tau_{FOV} \gg t_0$. This corresponds to a hard-to-find target. In that case

$$P_s = 1 - \left(1 - \frac{t_0}{\tau_{FOV}} + \cdots\right) \approx \frac{t_0}{\tau_{FOV}} \quad \text{when } \tau_{FOV} \gg t_0 \tag{15.26a}$$

so under these conditions

$$\tau_{FOR} = \frac{2 - \dfrac{t_0}{\tau_{FOV}}}{2\dfrac{t_0}{\tau_{FOV}}}(n-1)t_0 + \tau_{FOV} = n\tau_{FOV} + \frac{1}{2}(n-1)t_0$$

$$\approx n\tau_{FOV} \text{ when } \tau_{FOV} \gg t_0 \tag{15.26b}$$

which shows that when $\tau_{FOV} \gg t_0$ (15.25) reduces to (15.24), which was already shown to be reasonable for hard-to-find targets. In the case of an obvious target, we expect $\tau_{FOV} \ll t_0$. In that case $P_s \approx 1$ and (15.25) becomes

$$\tau_{FOR} = \frac{1}{2}(n-1)t_0 + \tau_{FOV}$$

$$\approx \frac{n}{2}t_0 \quad \text{when } \tau_{FOV} \ll t_0 \text{ and } n \gg 1 \tag{15.26c}$$

Equation (15.26c) is just what is expected for this case. For an obvious target and a large number of fields of view, the target on the average will be found after examining half the fields of view and since each field of view is examined for t_0, (15.26c) is a reasonable estimate for the mean time to acquire a target for this case.

The parameter P_s defined in (15.25) has an important physical interpretation. It is the probability that an observer who eventually detects the target will detect the target after one complete sweep of the FOR. This follows from the realization that the empty fields of view contribute nothing to the probability of target acquisition and (15.3a) asserts that an observer (we only consider those who eventually detect the target) who examines the FOV for a time t_0 will have a probability P_s of detecting the target. The s subscript in P_s stands for *success* in finding the target in any single sweep of the FOR.

The observation that (15.25) is dimensionally correct and behaves reasonably for extreme values of the parameters lends credibility to (15.25). This equation is graphically illustrated in Figure 15.14.

The construction of the graph illustrated in Figure 15.14 is facilitated by expressing (15.25) in terms of two dimensionless parameters

$$\pi_1 \equiv \frac{\tau_{FOR}}{t_0}; \quad \pi_2 \equiv \frac{\tau_{FOV}}{t_0} \tag{15.26d}$$

Then (15.25) is expressed by

$$\pi_1 = \frac{\left(3 - e^{-\frac{1}{\pi_2}}\right)}{2\left(1 - e^{-\frac{1}{\pi_2}}\right)}(n-1) + \pi_2 \tag{15.26e}$$

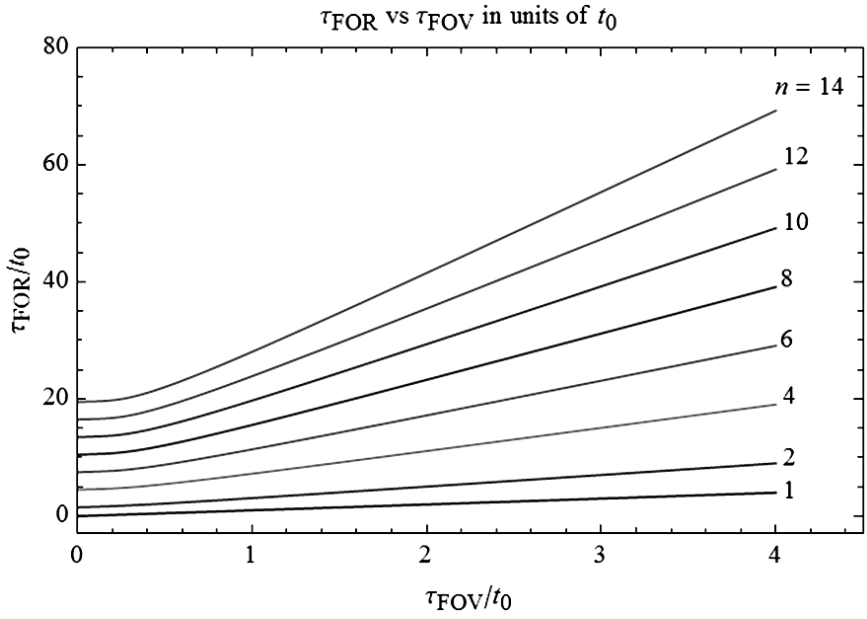


Figure 15.14 The relationship between τ_{FOR} and τ_{FOV} as a function of *n*, the number of fields of view in the FOR.

Figure 15.14 was constructed using (15.26e).

A mathematical description of how $P(t)$ grows during a FOR search using the time-unlimited search model described in Figure 15.13 is delayed until Section 15.9. The result is complicated. However, if one assumes the target has an equal probability of being in any one of the fields of view, then $P(t)$ can be approximated by (15.23) with τ_{FOR} given by (15.25). If one accepts the approximation that $P(t)$ is approximated by (15.23), then the same mathematics that led to (15.17) implies that in a FOR search limited by time t_L

$$P(t) = \begin{cases} P_{t_L} \dfrac{1 - e^{-t/\tau_{FOR}}}{1 - e^{-t_L/\tau_{FOR}}} & 0 \le t \le t_L \\ P_{t_L} & t_L < t \end{cases} \tag{15.27}$$

where P_{t_L} is the fraction of observers who detect the target in a FOR search that takes place for t between 0 and t_L. The mean time to find a target in a time-limited FOR search is given by (15.18) with τ_{FOV} replaced by τ_{FOR}:

$$< t > = \tau_{FOR} - \frac{t_L}{e^{t_L/\tau_{FOR}} - 1} \tag{15.28}$$

There is a need for metrics that characterize the effectiveness of FOR searches. One metric, which depends on time, is the number of detected targets $N_{det}(t)$ in the FOR. Searching is effective if $N_{det}(t)$ is a rapidly rising function and has a high value at time t_L when the observers stop searching. Another metric is $N_{det}(t_L)/N_{tgt}$, where N_{tgt} is the number of targets in the FOR and t_L is the amount of time the observers are allowed to search for the target.

15.7 Multiple Observers, Single Sensor, Unlimited Time, and Shared Knowledge

In this section we generalize the FOV search equation described by (15.1a), which describes the search performance of a single observer or an ensemble of observers in which *none of the observers communicate with one another*. In this section we assume there are *n identical* observers who can communicate with one another and thus the detection time is determined by the time at which the *first* observer detects the target. By *identical* we mean that each observer has an identical probability P_∞ of detecting the target, and the time constant with which each observer detects the target is τ. Just as there are differences in the times well-trained athletes will run a race, there are differences in the values for P_∞ and τ for well-trained soldiers searching for a target. Like the frictionless surfaces in mechanics, identical observers are a fiction created to get approximate answers to the questions asked in this section. Besides assuming identical observers, we also assume the observers search independently.

Assuming n identical independent observers, let $\mathbb{T}_{\min}(n)$ denote the detection time. In any experiment $\mathbb{T}_{\min}(n)$ is a random variable. We show in Appendix 15B that

$$< \mathbb{T}_{\min}(n) > = \tau \frac{\sum_{m=1}^{n} \frac{C(n,m)}{m} P_\infty^m (1-P_\infty)^{n-m}}{1-(1-P_\infty)^n} \tag{15.29}$$

where $< \mathbb{T}_{\min}(n) >$ is the expected value of $\mathbb{T}_{\min}(n)$, n is the number of observers, and $C(n, m)$ is defined by

$$C(n,m) \equiv \frac{n!}{m!(n-m)!} \tag{15.30}$$

The symbol $C(n, m)$ can be interpreted as the number of combinations for n things taken m at a time. Equation (15.29) depends on P^∞, τ, and n in a complicated way that makes it difficult to intuitively see the validity of the result. Equation (15.29) is illustrated in Figure 15.15. The dots in Figure 15.15 are the results of simulations done with 10,000 draws, which were done to confirm or refute the validity of (15.29). Except for slight differences between the model and the simulation for the case of two and three observers when P_∞ is less than 0.4, the simulation and (15.29) agree to within the size of the dot used to represent the simulation. With a small number of observers and a low P_∞, the number of successes is relatively low, so under these conditions the simulation is less accurate than (15.29).

Note from Figure 15.15 that when P_∞ is near one and the number of observers is large, $< \mathbb{T}_{\min}(n) > / \tau \ll 1$. Thus, *the mean detection time for a large number of observers searching for a target, with an ability to communicate with one another, is much lower than the mean detection time for a single observer*. In Appendix 15B we show that

$$\begin{aligned} < \mathbb{T}_{\min}(n) > &= \frac{\tau}{n_d} \\ &= \frac{\tau}{n} \quad \text{when } P_\infty = 1 \end{aligned} \tag{15.31}$$

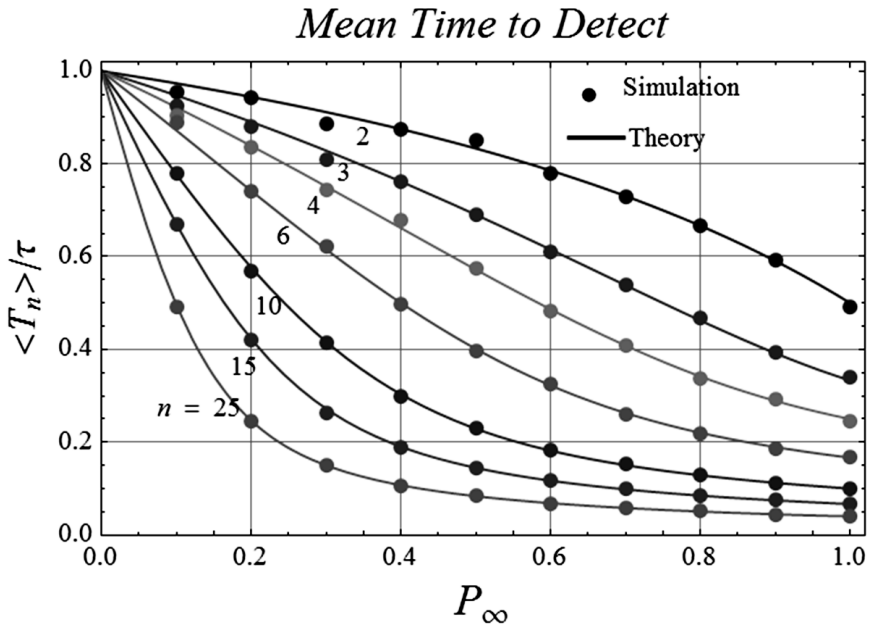


Figure 15.15 Graphical representation of (15.29). In this graph $\mathbb{T}_{\min}(n)$ is denoted by T_n.

where n_d is the number of observers who detect the target. Equation (15.31) predicts—unrealistically—that $< \mathbb{T}_{\min}(n) >$ approaches zero as n approaches infinity for easy-to-find targets for which P_∞ is one. The reason for this unrealistic result is that the model used to derive (15.29) assumed a zero delay time. We anticipate that a more realistic model would predict that $< \mathbb{T}_{\min}(n) >$ approaches t_d as n approaches infinity when P_∞ is near one.

Although (15.29) was derived for FOV searches, it is expected to be valid for FOR searches as long as (15.23) is valid. In Ref. [20], a FOR perception experiment, albeit not intended to investigate the validity of (15.29), was reanalyzed to determine if the results support (15.29). A comparison between (15.29) and the perception experiment is shown in Figure 15.16, where the dots and the error bars represent results from the perception experiment and the solid line is the prediction of (15.29). It was not possible to estimate the measurement uncertainty at $P_\infty = 0.3$, and there is large experimental uncertainty in the measurement done at $P_\infty = 0.6$. Although there is remarkable agreement between the perception experiment and the model, further experiments are needed to validate (15.29).

Suppose n identical indpendent observers each characterized by P_∞ are simultaneously searching a FOV. It is assumed that the target and the sensor are stationary and the observers are in communication with one another so the target is detected when any one observer detects the target. Appendix 15B shows that under these conditions the probability of detection P_{Det} is given by

$$P_{Det}(n, P_\infty) = 1 - (1 - P_\infty)^n \tag{15.32}$$

Equation (15.32) is illustrated graphically in Figure 15.17, which compares (15.32) with the results of a computer simulation done with 10,000 draws assuming

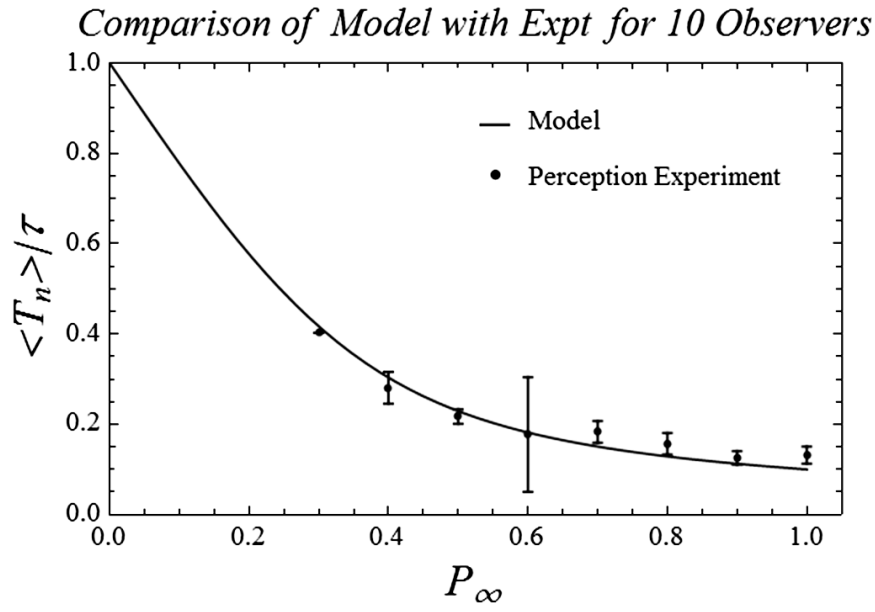


Figure 15.16 Comparison of (15.29) with perception experiment results. In this graph $\mathbb{T}_{min}(n)$ is denoted by T_n.

identical independent observers. The agreement between the simulated results and (15.32) suggests that (15.32) is valid for identical independent observers. However, we are not aware of a perception test that has been done to investigate the validity of (15.32) with real sensors, targets, and backgrounds. It would be remarkable if the results indicated by Figure 15.17 are true. The graph implies that in a perception experiment with 10 observers where the probability of a single observer detecting

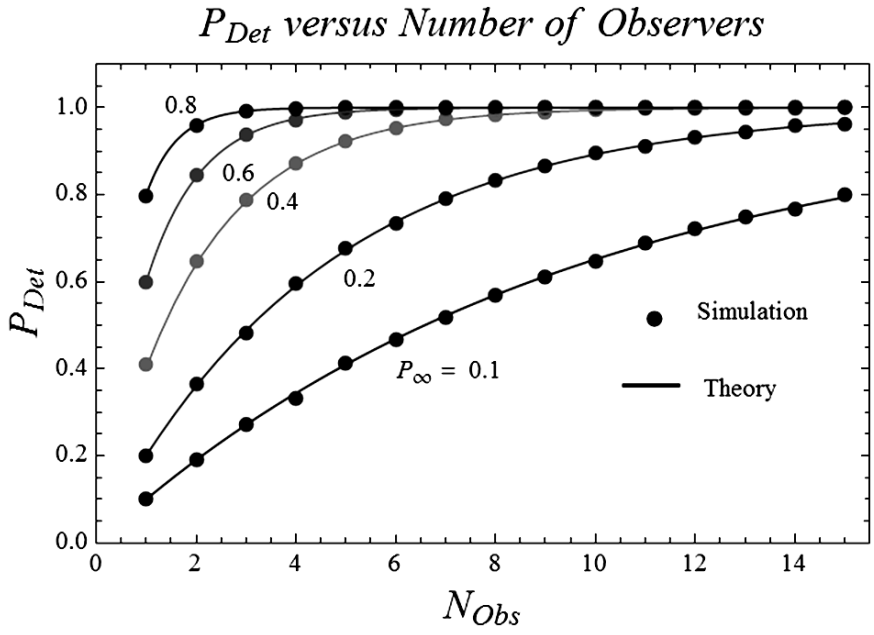


Figure 15.17 Comparison of (15.32) with simulated results. In this graph n is denoted by N_{Obs}.

the target is 0.1, the probability that one or more observers will detect the target is 0.65—a six- or sevenfold increase in detection probability from that of a single observer. Even if (15.32) does not accurately predict detection probabilities when multiple observers search a display produced by a single sensor, it suggests remarkable improvements in detection probabilities are possible when a single observer is replaced by multiple observers. We are not aware of experiments that have been done to either confirm or refute (15.32).

15.8 Independent Search with Two Sensors, Unlimited Time, and Shared Knowledge

In this section we consider FOV searching as described by (15.1a) or FOR searching as described by (15.23). We assume there are two observers, each with his or her own imaging sensor. Each sensor is characterized by an individual P_∞ and τ that describe the probability each observer has for eventually finding the target and the speed with which the observers find the target. Here it is assumed the targets and sensors are stationary. The two sensors do not have to be colocated or use the same radiation. Because the observers can communicate with one another, a target is said to be detected when the first of the two observers detects the target. In this section we calculate the mean time to detect the target $< \mathbb{T}_{\min} >$. Note that $\mathbb{T}_{\min}$ is a random variable and so we also calculate the probability density function $f_{\mathbb{T}_{\min}}$ that describes $\mathbb{T}_{\min}$.

We characterize imaging sensors 1 and 2 with the parameters $P_{\infty 1}$, τ_1, $P_{\infty 2}$, and τ_2. We have not put a FOV or FOR subscript on any of these variables because although the theory is developed with FOV searching in mind, it also applies to FOR searching providing (15.23) is an adequate approximation. Then[21]

$$< \mathbb{T}_{\min} > = \frac{P_{\infty 1}P_{\infty 2}}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}} \frac{\tau_1\tau_2}{\tau_1 + \tau_2} + \frac{P_{\infty 1}(1 - P_{\infty 2})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}}\tau_1 + \frac{P_{\infty 2}(1 - P_{\infty 1})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}}\tau_2 \tag{15.33}$$

A derivation of this result is given in Appendix 15B at the end of this chapter. This equation is a function of four variables and is difficult to visualize. However, the reader can easily confirm the reasonableness of (15.33). Observe that both the

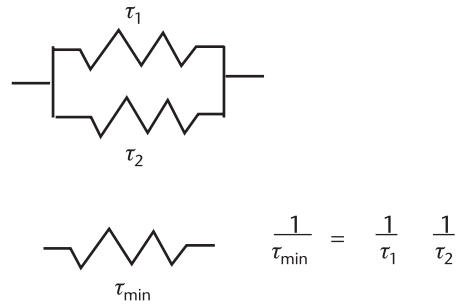


Figure 15.18 When $P_{\infty 1} = P_{\infty 2} = 1$, then an estimate for $\tau_{\min}$ is obtained by considering $\tau_{\min}$ as the effective resistance of two resistors in parallel where the values of the resistors are τ_1 and τ_2.

left- and right-hand sides of the equation are dimensionally equal to a time. Now consider the following limiting cases:

When $P_{\infty 2}$ is zero, then $< \mathbb{T}_{\min} > = \tau_1$; and when $P_{\infty 1}$ is zero, then $< \mathbb{T}_{\min} > = \tau_2$
When $P_{\infty 1} = P_{\infty 2} = 1$, then $< \mathbb{T}_{\min} > = (\tau_2 \tau_1)/(\tau_1 + \tau_2)$.

The first case is reasonable. If $P_{\infty 2} = 0$, then we would expect the mean time to detect a target would be that of sensor 1 and from (15.5a) that is τ_1. A similar argument can be made if $P_{\infty 1} = 0$. The second case is also reasonable. If $P_{\infty 1} = P_{\infty 2}$, then both sensors contribute to detecting the target and we would expect $< \mathbb{T}_{\min} >$ to be less than the smaller of and τ_1 and τ_2. The expression for $< \mathbb{T}_{\min} >$ given by (15.29) under these conditions is similar to that for the effective resistance of two resistors in parallel. Of course, under these conditions the effective resistance is less than either resistor and so the second case implies that $< \mathbb{T}_{\min} >$ is less than the smaller of τ_1 or τ_2. If we specialize even further to the case where $P_{\infty 1} = P_{\infty 2} = 1$ and $\tau_1 = \tau_2 = \tau$, then (15.29) implies that $< \mathbb{T}_{\min} > = \tau/2$. We would expect that if the target is found with two equally capable sensors and the target would eventually be found with either sensor, then the mean time to detect the target would be half the time needed to detect the target with a single sensor. Even without following the derivation, we conclude from the dimensional consistency of the result and the limiting cases that (15.33) is a reasonable result. Note that in deriving (15.33) it was assumed that t_d for both sensor–observer systems is negligible and can be neglected.

To describe the pdf associated with the random variable $\mathbb{T}_{\min}$, it is convenient to introduce a notation for the exponential pdf:

$$\mathbb{E}(t;\tau) \equiv \frac{1}{\tau} e^{-t/\tau}, \quad 0 \le t < \infty \tag{15.34}$$

Then the pdf for $\mathbb{T}_{\min}$ is given by

$$\begin{aligned} f_{\mathbb{T}_{\min}} = {} & \frac{P_{\infty 1} P_{\infty 2}}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1} P_{\infty 2}} \mathbb{E}\left(t; \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}\right) + \frac{P_{\infty 1}(1 - P_{\infty 2})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1} P_{\infty 2}} \mathbb{E}(t, \tau_1) \\ & + \frac{P_{\infty 2}(1 - P_{\infty 1})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1} P_{\infty 2}} \mathbb{E}(t, \tau_2) \end{aligned} \tag{15.35}$$

This equation is a function of four variables and is difficult to visualize. It is derived in Appendix 15B.4, where it is shown that it has the proper limiting behavior.

15.9 Time-Dependent Search Parameters Search Model

In this section we allow P_∞ and τ to be continuous functions of time. Some of the many scenarios in which P_∞ and τ are functions of time are described here:

1. Consider FOV or FOR searches where either the target or the sensor or both are moving so that the sensor–target range is changing. From the NV-IPM

model we know that under these conditions P_∞ will be a continuous function of time. Then from (15.11a) through (15.11i) we know that τ will also depend on time.

2. Consider a FOV or FOR search where either the target or the sensor or both are moving perpendicular to the sensor–target line of sight. Under these conditions the range is not changing but because the background against which the target is seen is changing, the target contrast is changing. Because of the changing contrast, under these conditions P_∞ and τ are continuous functions of time.
3. When the sensor changes from the wide to the narrow FOV, P_∞ and τ change.
4. When the target goes in and out of the field of view because it is occluded by terrain when either the target or sensor is moving, then P_∞ becomes zero and τ is not defined. The fact that τ is not defined when P_∞ is zero is unimportant since the calculation does not depend on the value of τ under this condition. When there is smoke, dust, or changing atmospheric conditions P_∞ and τ change.
5. In a FOR search, because of sensor slewing, there are moments when the target is in the FOV and moments when the target is out of the FOV. If the target is detectable when it is in the FOV, then during this time it can be characterized by P_∞ and τ. When the target is out of the FOV then $P_\infty = 0$ and τ is not defined.

Note that searching for improvised explosive devices from a vehicle moving along a road corresponds to case 1, so under these conditions P_∞ and τ depend on time.

In this section we describe the assumptions of searches that have time-dependent search parameters (TDSP) and present pertinent equations that describe the search process under these conditions.

The method used to get an analytical solution for the probability of target acquisition when P_∞ and τ are functions of time is involved. For that reason we outline the approach used to solve the problem. First, based on physical insight, the necessary properties of the solution are enumerated. We then assume a finite number of intervals with each interval small enough so that during each interval P_∞ and τ can be approximated as constant. Remarkably, satisfying the enumerated properties determines the solution. In particular, the condition that the acquisition probability be a monotonic, causal, and nondecreasing function of time enables the solution in the nth interval to be related to the solution in the $n-1$th interval. An algebraic expression is generated that describes how target acquisition probability grows for a relatively simple case where $P_{\infty i}$ and τ_i are specified by $i = 1, 2, 3$. Even in this simple case, the solution is complicated and illustrates the difficulty of writing explicit algebraic expressions when there are a large number of intervals. However, the algebraic expression provides guidance for implementing an iterative *Mathematica* procedure for calculating target acquisition probability with the TDSP model. The *Mathematic* a code is then used to investigate the validity of FOR search model (15.23).

In describing TDSP search theory, it is convenient to modify the notation we have used up to this point. Equation (15.1a) is written

$$P_r(t) = P_\infty(1 - e^{-t/\tau}) \quad 0 \le t < \infty \tag{15.36}$$

Here $P_r(t)$ is the *reduced* distribution function for acquiring the target. This is distinguished from the *cumulative* distribution function [see (15.3a)] associated with (15.36):

$$P_c(t) = (1 - e^{-t/\tau}), \quad 0 \le t < \infty \tag{15.37}$$

Equations (15.36) and (15.37) are the starting points for the development of the theory, and we see that these equations have built into them the assumptions of unlimited time to search and zero delay time t_d. The CDF always has an asymptotic value of one, whereas the reduced distribution function has an asymptotic value equal to P_∞ that is always less than or equal to one. Equation (15.36) gives the probability of detecting a target at a fixed sensor–target range in a time less than or equal to t for *all observers* taking part in a perception experiment, whereas (15.37) gives the probability of detecting a target at a fixed sensor–target range in a time less than or equal to t for *those observers in the perception experiment who eventually detect the target*.

Other quantities of interest are the probability density function $p_c(t)$ and the reduced probability density function $p_r(t)$. These are related to $P_c(t)$ and $P_r(t)$:

$$p_c(t) = \frac{dP_c(t)}{dt} \tag{15.38}$$

$$p_r(t) = \frac{dP_r(t)}{dt} \tag{15.39}$$

We demand that the equations we develop satisfy all the properties listed in Table 15.1. Properties 1, 2, and 3 are required of any CDF and, since $P_r(t)$ is related to a CDF by a constant factor, these properties must be satisfied. Physically, we expect $P_r(t)$ to be a continuous function of the search parameters $P_\infty(t)$ and $\tau(t)$ and this implies property 4. Property 5 asserts that $P_r(t)$ never gets greater than the maximum $P_{\infty i}$. Intuitively we believe this requirement makes sense and expect that this requirement is very nearly true. With regard to property 6, as $P_r(t)$ grows with increasing time, it is reasonable to allow the way $P_r(t)$ grows to depend on earlier and current values for P_∞ but it cannot depend on future P_∞ values. Of course, our model must agree with experiment as indicated by property 7. As we will subsequently see, properties 1 through 6 impose a particular solution.

Table 15.1 Required Properties of $P_r(t)$

No	*Property*
1	$P_r(t)$ must be continuous.
2	$0 \le P_r(t) \le 1$ for all t.
3	$P_r(t)$ must be nondecreasing.
4	If $P_\infty(t)$ and $\tau(t)$ are nearly constant, then $P_r(t)$ and the result of (15.36) are nearly the same.
5	$P_r(t) < max(P_{\infty 1}, P_{\infty 2}, \ldots, P_{\infty n})$ for all $t > 0$.
6	$P_r(t)$ must be causal.
7	$P_r(t)$ and $P_c(t)$ must agree with experiment.

Understanding (15.40), given below, is key to understanding the solution to the problem discussed in this section. We introduce the notation T_i^- to indicate the time $T - \epsilon$ where ϵ is a vanishingly small positive quantity. So T_i^- corresponds to a time infinitesimally smaller than T_i. Suppose $P_r(T_i^-) = \alpha$ and that for all time greater than or equal to T_i, a search is characterized by search parameters $P_{\infty i}$ and τ_i. How does the probability of target acquisition grow for $t \geq T_i$? Let $P_i(t)$ denote the value of $P_r(t)$ for time greater than or equal to T_i. Our solution depends on the relative size of $P_{\infty i}$ and α:

$$P_i(t) = \begin{cases} \alpha + (P_{\infty i} - \alpha)\left(1 - e^{\frac{-t-T_i}{\tau_i}}\right) & t \geq T_i \text{ AND } P_{\infty i} > \alpha \\ \alpha + (0)\left(1 - e^{\frac{-t-T_i}{\tau_i}}\right) & t \geq T_i \text{ AND } P_{\infty i} \leq \alpha \end{cases} \tag{15.40}$$

Equation (15.40) has the following properties:

1. Regardless of the relative size of $P_{\infty i}$ and α, $P_i(T_i) = \alpha$. This guarantees that $P_r(t)$ is a continuous function as it must be.
2. When $P_{\infty i} > \alpha$, then the asymptotic value of $P_i(t)$ as t gets arbitrarily large is $P_{\infty i}$ as expected.
3. When $P_{\infty i} > \alpha$, then $P_i(t)$ approaches $P_{\infty i}$ with a time constant τ_i as expected.
4. When $P_{\infty i} \leq \alpha$, then for $t > T_i$, $P_i(t)$ has the constant value α as expected.

Property 1 guarantees that $P_r(t)$ is a continuous function of t, as it must be. Properties 2 and 3 express mathematically what is meant by a time interval characterized with search parameters $P_{\infty i}$ and τ_i. Property 4 is required by properties 3, 5, and 6 in Table 15.1. Using the technique expressed by (15.40) and the conditions required in Table 15.1 we find the solution for $P_r(t)$ by repeatedly using (15.40) to calculate how the probability of target acquisition grows in the nth interval from knowledge of how the target acquisition grew in the $n - 1$th interval.

It is convenient to express our solution in terms of the *Heaviside theta* function, which is defined in *Mathematica* as folows:

$$\vartheta(t) \equiv \begin{cases} 0 & t < 0 \\ 1 & t > 0 \\ \text{undefined} & t = 0 \end{cases} \tag{15.41}$$

To express our answer we will find a generalization of the unit step function U_g useful (the g subscript is short for *generalization*):

$$U_g(t) \equiv \begin{cases} t\vartheta(t) & t \neq 0 \\ 0 & t = 0 \end{cases} \tag{15.42}$$

Although $P_\infty(t)$ and $\tau(t)$ are functions of time, they can be approximated as constants in a small enough time interval. To develop our model, we start by assuming

three times defined by $T_1 = 0$, T_2 and T_3 that are the boundaries of four time intervals. Later we will generalize to n intervals where n can be as large as desired. Let $P_{\infty 1}$, $P_{\infty 2}$, $P_{\infty 3}$, τ_1, τ_2, and τ_3 denote the pertinent search parameters in each interval. We use the convention that the i subscripts on $P_{\infty i}$ and τ_i correspond to the beginning of the time period. Thus, $P_{\infty 2}$ and τ_2 correspond to search parameters pertinent to the time period between T_2 and T_3.

Using the technique of (15.40) and the first six properties specified in Table 15.1, we can determine the solution for $P_r(t)$ with four time intervals:

$$P_r(t) = \begin{cases} 0 & -\infty < t < 0 \\ P_{\infty 1}\left(1 - Exp\left[-\dfrac{t - T_1}{\tau_1}\right]\right) & T_1 \le t < T_2 \\ P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right) + U_g\left[P_{\infty 2} - P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right)\right]\left(1 - Exp\left[-\dfrac{t - T_2}{\tau_2}\right]\right) & T_2 \le t < T_3 \\ P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right) + U_g\left[P_{\infty 2} - P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right)\right]\left(1 - Exp\left[-\dfrac{T_3 - T_2}{\tau_2}\right]\right) & \\ \quad + U_g\left[P_{\infty 3} - P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right) + U_g\left[P_{\infty 2} - P_{\infty 1}\left(1 - Exp\left[-\dfrac{T_2}{\tau_1}\right]\right)\right]\right] & \\ \quad \left(1 - Exp\left[-\dfrac{T_3 - T_2}{\tau_2}\right]\right)\Bigg] \times \left(1 - Exp\left[-\dfrac{t - T_3}{\tau_3}\right]\right) & T_3 \le t < \infty \end{cases} \tag{15.43}$$

In this equation the function U_g switches appropriately between the top and bottom right-hand sides of (15.40). Equation (15.43) is derived by repeatedly using (15.40) and can be understood by looking at it carefully. In the time interval:

$-\infty < t < 0$, $P_r(t)$ is zero because the search has not yet begun.

$T_1 \le t < T_2$ conditions 1, 4, 6, and 7 imply $P_r(t) = P_{\infty 1}\left(1 - exp\left[-\dfrac{t}{\tau_1}\right]\right)$

$T_2 \le t \le T_3$ conditions 1, 5, and 6 imply

$$P_r(t) = P_{\infty 1}\left(1 - exp\left[-\frac{T_2}{\tau_1}\right]\right) + U_g\left[P_{\infty 2} - P_{\infty 1}\left(1 - exp\left[-\frac{T_2}{\tau_1}\right]\right)\right]\left(1 - exp\left[-\frac{t - T_2}{\tau_2}\right]\right)$$

The development of $P_r(t)$ for $t < T_2$ is obvious. In the interval $T_2 \le t < T_3$, the term $P_{\infty 1}\left(1 - exp\left[-\dfrac{T_2}{T_1}\right]\right)$ corresponds to α in (15.40), and $U_g\left[P_{\infty 2} - P_{\infty 1}\left(1 - exp\left[-\dfrac{T_2}{T_1}\right]\right)\right]$ corresponds to $P_{\infty i} - \alpha$ in (15.40) or to zero depending on the relative size of $P_{\infty 2}$

and α. Note that the second term vanishes when $t = T_2$, which implies the first term has to have the value given so the value at $t = T_2$ is the same in this interval as it was in the previous interval. The factor $\left(1 - exp\left[-\frac{t - T_2}{\tau_2}\right]\right)$ ensures that $P_r(t)$ will asymptote to the value $P_{\infty 2}$ with time constant τ_2 providing it is not interrupted by future $P_{\infty 3}$, τ_3 inputs.

In the time interval $T_3 \le t < \infty$ the pattern established in the preceding interval repeats itself:

- The first line for this interval is needed to guarantee continuity of $P_r(t)$.
- The argument of the outer U_g term is needed to ensure that in the absence of future changes in P_∞ the expression will asymptote to $P_{\infty 3}$ providing $P_{\infty 3}$ is larger than the value of $P_r(t)$ at the end of the last interval but will stay constant if this is not the case.
- The last factor guarantees that if $P_{\infty 3}$ is larger than the value of $P_r(t)$ at the end of the last interval, then $P_r(t)$ will approach the asymptotic value $P_{\infty 3}$ with time constant τ_3.

Although we can satisfy the properties specified in Table 15.1 with the approach of (15.43), it is apparent that with a large number of intervals the expression will get exceedingly complex. An iterative approach captures the idea of (15.43) and is also amenable to a large number of intervals.

In the iterative approach we expand $P_r(t)$ as a sum of $P_i(t)$ where $P_i(t)$ has a finite value in the interval $T_i < t \le T_{i+1}$ and is zero outside that interval. To accomplish this, we introduce a function $U(t, T_i, T_{i+1})$:

$$U(t, T_i, T_{i+1}) \equiv \begin{cases} 1 & T_i < t \le T_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{15.44}$$

The time dependence in (15.43) for each time interval has the functional form. $1 - exp\left[-\frac{t - T_i}{\tau_i}\right]$. The iterative expression for $P_r(t)$ will be more compact and easier to write if we introduce a function $W(t, T, \tau)$:

$$W(t, T, \tau) \equiv 1 - \exp\left[-\frac{t - T}{\tau}\right] \tag{15.45}$$

Identify $P_i(t)$ with the ith time interval in (15.43). Examining (15.43) enables us to write an iterative expression for $P_i(t)$:

$$P_0(t) \equiv 0 \tag{15.46a}$$

$$P_i(t) = \{P_{i-1}(T_i) + U_g[P_{\infty i} - P_{i-1}(T_i)]W(t, T_i, \tau_i)\}U(t, T_i, T_{i+1}) \tag{15.46b}$$

Equation (15.46b) enables us to calculate $P_1(t)$ in terms of $P_0(t)$. Using (15.46b) again, we can calculate $P_2(t)$ in terms of $P_1(t)$. Using (15.46b) repeatedly, $P_i(t)$ can be written with i equal to any integer n in terms of preceeding values that have been calculated.

In (15.46b), $P_i(t)$ is defined in terms of P_{i-1}, T_i, and τ_i. These quantities need to be defined. The notation T_i denotes the time the ith interval begins. Thus, T_1 denotes the time the first interval begins, which is zero. With n values input for T_i, $P_{\infty i}$, and τ_i, note from the definition of $P_i(t)$ that it is also necessary to specify T_{n+1}, which is considered infinite; this allows $P_r(t)$ to be calculated anywhere in the interval $0 \le t < \infty$. The time constant describes how $P_i(t)$ grows in the ith interval, and if $P_{\infty i}$ is greater than $P_{i-1}(T_i)$, then $P_{\infty i}$ is the asymptotic value that $P_r(t)$ would approach if T_{i+1} were infinite. Note from the definition of $P_i(t)$ that when $i = 1$, we need to specify $P_{\infty 0}$ to be zero; in the function U_g there is a call for $P_{i-1}(T_i)$ and this requires that we set $P_{\infty 0}$ to zero. Since P_i is defined in terms of P_{i-1}, setting $P_0(t)$ to a definite value causes the process to stop.

The input parameters are T_i, $P_{\infty i}$, and τ_i for $i = 1,\ldots,n$. As indicated in (15.46), we specify $T_1 = 0$, $T_{n+1} = \infty$, and $P_{\infty 0} = 0$. With these boundary conditions (15.46) works as intended.

$$\{T_1 = 0, T_2, \ldots, T_n, T_{n+1} = \infty\} \tag{15.47a}$$

$$\{P_{\infty 0} = 0, P_{\infty 1}, P_{\infty 2}, \ldots, P_{\infty n}\} \tag{15.47b}$$

$$\{\tau_1, \tau_2, \ldots, \tau_n\} \tag{15.47c}$$

In the iterative approach, $P_r(t)$ is the sum of the $P_i(t)$ values:

$$P_r(t) = \sum_{i=1}^{n} P_i(t) \tag{15.48}$$

The cumulative distribution function $P_c(t)$ has the property that it is proportional to $P_r(t)$ but with the property that $P_c(\infty) = 1$. It is defined by

$$P_c(t) = \frac{P_r(t)}{P_r(\infty)} \tag{15.49}$$

The reduced probability density function $p_r(t)$ for target detection is given by

$$p_r(t) = \frac{dP_r(t)}{dt} \tag{15.50}$$

The probability density function $p_c(t)$ for target detection is given by

$$p_c(t) = \frac{dP_c(t)}{dt} \tag{15.51}$$

Although $P_{\infty i}$ and τ_i are input as lists that represent the search parameters in an interval, it is convenient to represent them as step functions of time. The functions $P_\infty(t)$ and $\tau(t)$ express the input lists $\{P_{\infty 1}, P_{\infty 2}, \ldots, P_{\infty n}\}$ and $\{\tau_1, \tau_2, \ldots, \tau_n\}$ as continuous functions of time.

$$P_\infty(t) = \sum_{i=1}^{n} P_{\infty i}\, U(t, T_i, T_{i+1}) \tag{15.52}$$

$$\tau(t) = \sum_{i=1}^{n} \tau_i \, U(t, T_i, T_{i+1}) \tag{15.53}$$

Mathematica code that implements the equations of this section is given in Appendix 15C.

Property 4 in Table 15.1 requires that the model developed in this section agree with (15.1a) when the search parameters are constant. The solid curve in Figure 15.19 was calculated using the code given in Appendix 15C using $\tau = 1$ sec and $P_\infty = 0.8$ and the points were calculated using (15.1a). This figure demonstrates that the TDSP search model solution satisfies property 4 in Table 1. This is expected since the TDSP search model is designed to satisfy (15.1a) exactly when the search parameters are constant.

The time-varying search parameter model will now be used to confirm the validity of the FOR search model (15.23) in which τ_{FOR} is given by (15.26b). To do that, we consider the particular case where the target is described by the search parameters $P_\infty = 0.8$ and $\tau_{FOV} = 4$ sec. In the case under consideration, seven fields of view are required to cover the FOR and this is done in a step-stare approach with $t_0 = 2$ sec; that is, each FOV is examined for 2 sec before moving on to the next FOV.

Using (15.25) we determine the values for P_s and τ_{FOR} : $P_s = 0.393$ and $\tau_{FOR} = 28.49$ sec. If the target is in the second FOV visited in the step-stare FOR search, then the probability of acquiring the target as a function of time is found using the code given in Appendix 15C and is illustrated in Figure 15.20.

In Figure 15.20, the places where the probability of target acquisition remains constant correspond to times in the step-stare search when the target is not in the FOV. In the derivation of the FOR search model, it was assumed that the target has an equal probability of being in each FOV in the FOR. As illustrated in Figure 15.20, if the target is in the second FOV, then it is found quicker than predicted by (15.23) because it is visited earlier than was modeled.

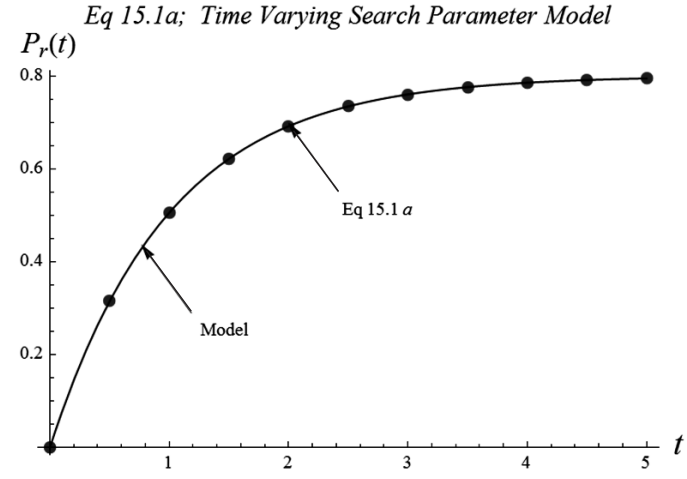


Figure 15.19 Comparison of TDSP search model with (15.1a).

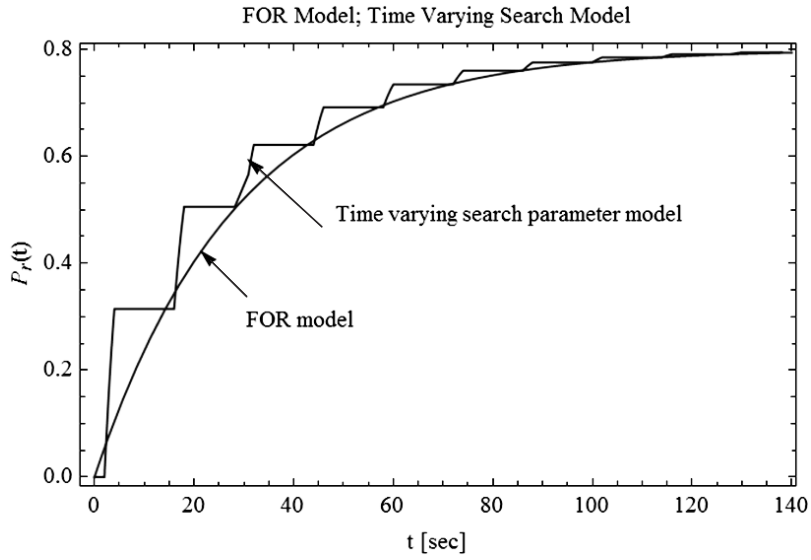


Figure 15.20 Comparison of the FOR search model of (15.23) with the TDSP search model when the target is in the second FOV.

Figure 15.21 shows how the probability of target acquisition grows during a step-stare FOR search when the target is in a particular FOV in the FOR. In Figures 15.20 through 15.23, we model step-stare FOR searches with $P_\infty = 0.8$, $\tau = 4$, $t_0 = 2$, and $n = 7$. In executing the code P_∞ and τ were sampled at a rate of once per second. Figure 15.21 illustrates how much quicker a target will be found in step-stare search if it is visited early.

Figure 15.22 compares TDSP search where the target has equal probability of being in the first and seventh FOV. Figure 15.23 compares TDSP search where

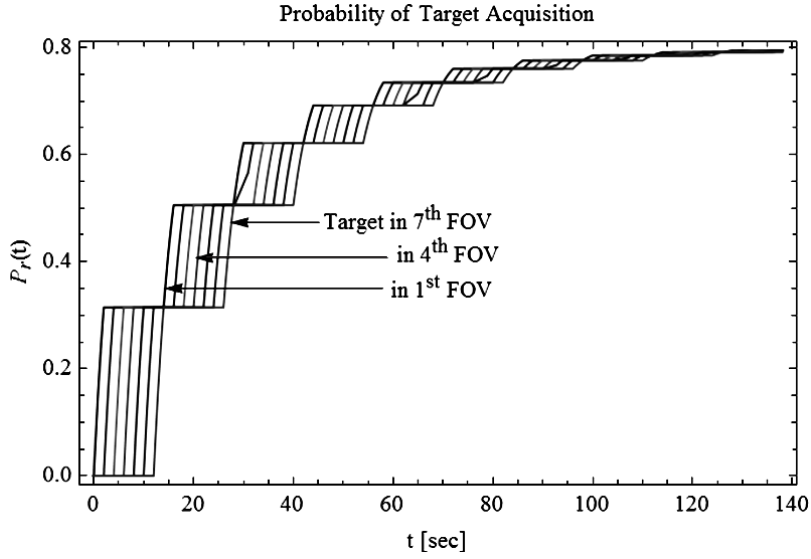


Figure 15.21 Probability of target acquisition in a step-stare search when the target is in a particular FOV in the FOR.

the target has equal probability of being in first, second, sixth, and seventh FOV. Comparison of Figures 15.22 and 15.23 suggests that if the target had an equal probability of being in each FOV, then the TDSP search model and the FOR model defined by (15.23) would be indistinguishable. This calculation has been done. On the scale of the graph, results obtained using (15.23) and (15.25) are indistinguishable from results calculated using the code in Appendix 15C.

An advantage of the FOR equations exhibited in Section 15.6 is they can be evaluated easily but require the assumption that the target has an equal probability of being in each FOV. An advantage of the results obtained using the code in Appendix 15C is it can be applied to cases where the target has an arbitrary probability of being in each FOV needed to cover the FOR.

The TDSP search model results shown here support the view that (15.23) and (15.25) are valid for the step-stare search approach providing the target has an equal probability of being in each FOV needed to cover the FOR.

15.10 Other Work

Neoclassical Search Model

The Neoclassical Search Model was developed by Jeffrey Nicoll[38,39] at the Institute for Defense Analysis (IDA). The model postulates three states: 1) attending to the target, 2) attending to other objects in the scene, and 3) wandering. The foveal point of attention follows a Markov process in making transitions between the three states. When in the "attending to the target state" there is some probability

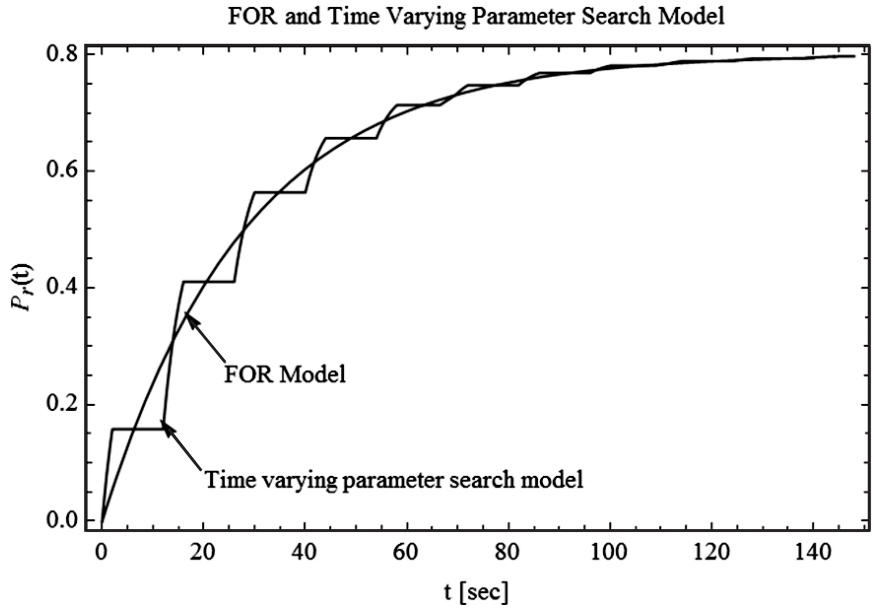


Figure 15.22 In the TDSP search model calculation, the target was assumed to be in the first and seventh FOV with equal probability. In the FOR search model, the target was assumed to be in each of the fields of view with equal probability.

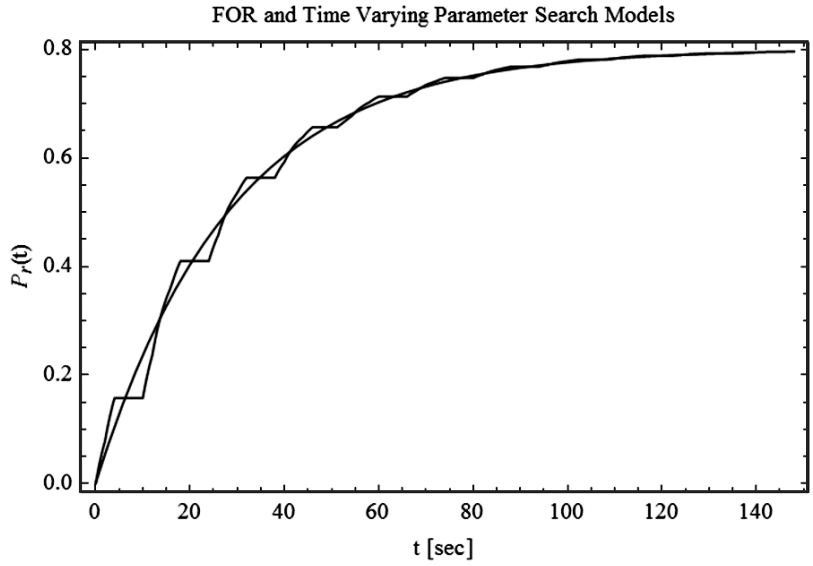


Figure 15.23 In the TDSP search model calculation, the target was assumed to be in the first, second, sixth, and seventh FOVs with equal probability. The FOR search model result assumes the target is in each FOV with equal probability. The TDSP search model and the FOR search model are indistinguishable on the scale of this graph when the target has an equal probability of being in every FOV needed to cover the FOR.

of making a transition to the detection state, which terminates the process. When in this state, however, it is also possible to transition into the "wandering" or "attending to some other object" state. With the exception of "detection," which terminates the process, the eye can make multiple visits to each of the three states.

The theory postulates the transition rates between the three states and initial probabilities of being in any one state. It is possible to write three linear differential equations that indicate how the probability of being in any one state varies with time, as well as a differential equation that describes how the detection probability grows with time. The model allows for the uniform treatment of multiple targets. The theory predicts that the probability of target acquisition P_d is given by

$$P_d(t) = \sum_{i=1}^{3} e_i(1 - e^{-\lambda_i t}) \tag{15.54}$$

Here the λ_i are eigenvalues associated with the three differential equations and the e_i are calculated from the initial probabilities of being in one of the three states.

Georgia Tech Vision Model

The Georgia Tech Vision (GTV) model is being developed to predict human visual search and target acquisition performance. The GTV model is an end-to-end simulation[7] of the human visual system. It accepts sequences of images that contain a target moving through a cluttered background at 30 Hz or greater. It segments the

images into objects and outputs predictions of the observer's search and detection performance. The model includes algorithms that describe dilation of the pupil, responses of the retinal receptors, luminance adaptation, and sensitivity to temporal modulation. Studies have been done[10] indicating that the GTV model accurately predicts human color discrimination and also produces detection probabilities that closely match Blackwell-type human threshold data.

Combat Simulations and War Games

War game modeling is an important application of search modeling. To develop tractable search models it is often necessary to assume identical independent observers. Perception tests show that observers differ in their ability to detect targets. The Target Acquisition Draw Methodology (TADM) developed by J. Silk[48,49] or variations of that methodology are used in war gaming methodologies. TADM[49] supplies observers with capabilities that are predominantly average with occasional observers who are extremely poor or extremely good. This is an improvement over the previous methodology, which gave observers a capability that was uniformly distributed between poor and good. TADM requires three random draws[9]: 1) an observer–target draw to account for the stochastic process, 2) an observer draw to account for the different capability levels of the observers, and 3) a target draw to account for the uncertainty in the target signature computation.

Acquire

ACQUIRE is the standard[1] model integrated into war games. Initially the ACQUIRE model was intended for determining how the probability of target acquisition would grow in conventional symmetric battlefield confrontations. Over a period measured in years, the ACQUIRE model has been enhanced and/or modified to improve its accuracy for specific uses. ACQUIRE-LC is used for low-contrast targets that are cleverly camouflaged. The ACQUIRE model has been modified to allow for time-limited search and is applicable to situations where the observer has a limited amount of time to find the target. Perception tests have been done that allow ACQUIRE to be used appropriately in an urban environment where the battlefield is asymmetric and where the objective is to distinguish terrorists from ordinary civilians. ACQUIRE has also been modified to account for moving targets.

Some things are not accounted for[32] in ACQUIRE:

1. Experiments indicate that a muzzle flash greatly increases the probability of detection.
2. The probability of detection increases greatly for a second target that is in proximity to a first target that has either been detected or is about to be detected.
3. Obscurants such as smoke are treated statically and a more proper treatment requires target acquisition with smoke to be treated dynamically.
4. The effect of the color, plain or patterned, of the target is not addressed.
5. Clutter in ACQUIRE is crudely handled by appropriately adjusting the V_{50} parameter.

Much effort[32] has been expended in an attempt to develop a more accurate target acquisition methodology, including several models based on computational vision models. However, none of the models examined[1] was found to be more accurate than ACQUIRE over the range of conditions, targets, and sensors of interest to army modelers. ACQUIRE is currently incorporated in the much more comprehensive NV-IPM model.

GIQE/IIRS/NIIRS

The General Image Quality Equation (GIQE), the Image Interpretability Image Rating Scale (IIRS), and the National Image Interpretability Image Rating Scale (NIIRS) tools are used by image analysts to interpret aerial imagery.[26,29,31] The NIIRS scale was created in 1994 and is a revised version of IIRS. The IIRS has a unique relationship between ratings that range between 0 and 9 and ground-resolved distances. Zero corresponds to imagery where only the largest objects can be interpreted, 1 corresponds to ground-resolved distances greater than 9 m, and an IIRS rating of 9 corresponds to a ground-resolved distance of less than 10 cm. NIIRS also has ratings that go from 0 to 9 and what can be done with different NIIRS imagery is given in tables.[26,29] For example, for an image with a NIIRS rating of 2, an image analyst could detect large hangars at airfields, and with a NIIRS rating of 3 an image analyst could detect a train. There are NIIRS ratings[29] for systems that use visible, infrared, and radar radiation.

15.11 Guide to the References

A working knowledge of probability theory is crucial for understanding search modeling. The probability books by Hsu,[28] Papoulis,[41] and Peebles[42] are recommended. Mauer[33] provides a good overview of search modeling. The Koopman[30] and Stone[50] references are for the reader interested in naval search applications from an operations research perspective. Rotman [43–46] discusses search theory from an army perspective.

15.12 Exercises

15.1 Show that the generalization of (15.5a) is (15.5c). Why is the mean time to acquire a target important?

15.2 Do computer simulations to verify the correctness of (15.6) and (15.9).

15.3 Do computer simulations to verify the correctness of (15.18) and (15.22).

15.4 Do computer simulations to verify the correctness of (15.29).

15.5 Do computer simulations to verify the correctness of (15.32). Design a perception experiment to investigate the validity of this equation.

15.6 Do a computer simulation to verify the correctness of (15.33). Generalize (15.33) to the case of three observers with three sensors each characterized by $(P_{\infty 1},\tau_1)$, $(P_{\infty 2},\tau_2)$, and $(P_{\infty 3},\tau_3)$.

15.7 Do computer simulations to verify the validity of (15.35).
15.8 Implement the code given in Appendix 15C in a computer language that you like. Then do the calculation illustrated in Figure 15.19.
15.9 Do your own calculation with the view of reproducing the step-stare results shown in Figure 15.20. Suppose the target is in the second FOV, but the seven different fields of view are visited randomly, staying at each FOV for a period of 2 sec. Reproduce Figure 15.20 for this case.
15.10 Do your own calculation to reproduce Figure 15.21.
15.11 Show that the last statement at the bottom of Figure 15.23 is true.
15.12 When there are just two fields of view covering the FOR, (15.23) is expected to be a relatively poor approximation in a step-stare search even if the targets have an equal probability of being in each FOV. Use the TDSP search model to investigate the validity of (15.23) under these conditions.
15.13 Show that $f_{\mathbb{T}_{\min}}(t)$ given by (15.35) has the following properties:
(a) $\int_0^\infty f_{\mathbb{T}_{\min}}(t)\,dt = 1$.
(b) $f_{\mathbb{T}_{\min}}(t) = \mathbb{E}(t;\tau_2)$ when $P_{\infty 1} = 0$ and $f_{\mathbb{T}_{\min}}(t) = \mathbb{E}(t;\tau_1)$, when $P_{\infty 2} = 0$.
(c) Equation (15.33) is implied by (15.35).

Property (a) shows that $f_{\mathbb{T}_{\min}}(t)$ satisfies an essential property of a pdf. Because the pdf of $\mathbb{T}_{\min}$ for sensor 2 operating alone is $\mathbb{E}(t;\tau_2)$ and the pdf of $\mathbb{T}_{\min}$ for sensor 1 operating alone is $\mathbb{E}(t;\tau_1)$, we expect $f_{\mathbb{T}_{\min}}(t)$ to satisfy property (b). The mean, the standard deviation, and all of the moments are implied by (15.35). Also (15.33) must be derivable from (15.35) if the two equations are consistent.

Appendix 15A: Time-Unlimited Field of Regard Search

In this appendix (15.25) is derived. The reader is referred to Section 15.6 for a description of the notation and the model described by (15.25).

We first list three mathematical results that will subsequently be needed:

$$\sum_{i=0}^{\infty} k^i = \frac{1}{1-k} \tag{15A.1a}$$

$$\sum_{i=1}^{\infty} ik^i = \frac{k}{(1-k)^2} \tag{15A.1b}$$

$$\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2} \tag{15A.1c}$$

Equation (15A.1a) is the well-known formula for the sum of a geometric series. Equation (15A.1b) is obtained from the preceding equation by differentiating it with respect to k. Equation (15A.1c) is the well-known formula for the sum of an arithmetic series.

Equation (15.13b), which gives the pdf for a time-limited search, can be expressed in terms of P_s:

$$p(t) = \frac{1}{\tau_{FOV} P_s} e^{-t/\tau_{FOV}} \quad 0 \le t < t_0 \tag{15A.2}$$

Let T denote the mean time to find the target given that the time-limited pdf of (15A.2) applies. Then

$$T = \int_0^{t_0} t \frac{1}{\tau_{FOV} P_s} e^{-\frac{t}{\tau_{FOV}}} dt \tag{15A.3}$$

$$= \frac{\tau_{FOV}}{P_s} - \frac{\tau_{FOV}}{P_s}\left(1 + \frac{t_0}{\tau_{FOV}}\right) e^{-t_0/\tau_{FOV}}$$

It is easy to show from (15A.3) that when $\tau_{FOV} \ll t_0$ then $T \approx \tau_{\mathrm{FOV}}$ and that when $\tau_{FOV} \gg t_0$ then $T \approx \tau_0/2$. These are reasonable results: 1) When $t_0 \gg \tau_{FOV}$, the observer has as much time as needed so (15.5a) applies. 2) When $t_0 \ll \tau_{FOV}$, the pdf of (15A.2) is approximately a uniform distribution between 0 and t_0 so the mean time to find the target is $t_0/2$.

Let t_1 denote the mean time to find the target for those observers who find the target if it is in the first FOV examined in the FOR. The observation that P_S is constant each time the FOR is surveyed allows us to assert the following:

1. The probability the target is found in the first pass of the FOR is P_S and the mean time to find the target given that the target was found in the first pass is T.
2. The probability the target is found in the second pass of the FOR is $P_S(1 - P_S)$ and the mean time to find the target given that the target is found in the second pass is $n\, t_0 + T$.
3. The probability the target is found in the third pass of the FOR is $P_S(1 - P_S)^2$ and the mean time to find the target given that the target is found in the third pass is $2n\, t_0 + T$.

Thus,

$$t_1 = P_s T + P_s(1 - P_s)(n\, t_0 + T) + P_s(1 - P_s)^2(2n\, t_0 + T) + \ldots \tag{15A.4a}$$

This equation can be simplified if we introduce the notation $k \equiv (1 - P_s)$. Then (15A.4a) becomes

$$t_1 = T\, P_s(1 + k + k^2 + \ldots) + t_0\, n\, P_s(k + 2k^2 + 3k^3 + \ldots) \tag{15A.4b}$$

Using (15A.1a) and (15A.1b), this can be written in the form

$$t_1 = T + t_0\, n\, P_s \frac{k}{(1 - k)^2} \tag{15A.4c}$$

Let t_2 denote the mean time to find the target for those observers who find the target if the target is in the second FOV. Then a calculation similar to the one done for t_1 yields the result

$$t_2 = t_0 + T + t_0\ n\ P_s \frac{k}{(1-k)^2} \tag{15A.5}$$

which reasonably asserts that if the target is in the second FOV then on the average it will be detected a time t_0 later. Let t_3 denote the mean time to find the target for those observers who find the target if the target is in the third FOV. Then a calculation similar to the one done for t_1 yields the result

$$t_3 = 2t_0 + T + t_0\ n\ P_s \frac{k}{(1-k)^2} \tag{15A.6}$$

The extension to the case where the target is in the mth FOV is apparent:

$$t_m = (m-1)t_0 + T + t_0\ n\ P_s \frac{k}{(1-k)^2} \tag{15A.7}$$

We assume the target has an equal probability of being in any field of view. Then the mean time to acquire the target is the arithmetic average of $t_1,t_2,\ldots,t_n$. Thus,

$$\tau_{FOR} = \frac{1}{n}(t_1 + t_2 + \ldots + t_n) \tag{15A.8a}$$

Using (15A.4c) and (15A.5) through (15A.7), we get

$$\tau_{FOR} = \frac{1}{n}(t_0(1 + 2 + \cdots + n-1)) + T + t_0\ n\ P_s \frac{k}{(1-k)^2} \tag{15A.8b}$$

Using (15A.1c) and the definition for k, this becomes

$$\begin{aligned} \tau_{FOR} &= \left\{ t_0 \frac{(n-1)}{2} \right\} + T + t_0\ n\ P_s \frac{k}{(1-k)^2} \\ &= T - \frac{t_0}{2} + t_0\ n \frac{(2-P_s)}{2P_s} \end{aligned} \tag{15A.8c}$$

We now show that (15A.8c) is identical to (15.25). Using the definition for P_s given in (15.25) enables us to express the exponential in (15A.3) in terms of P_s:

$$\begin{aligned} T &= \frac{\tau_{FOV}}{P_s} - \frac{\tau_{FOV}}{P_s}\left(1 + \frac{t_0}{\tau_{FOV}}\right)(1-P_s) \\ &= \tau_{FOV} + t_0 - \frac{t_0}{P_s} \end{aligned} \tag{15A.9}$$

where the second line is a mathematical simplification of the first line. Using this result, (15A.8c) becomes

$$\begin{aligned} \tau_{FOR} &= \left(\tau_{FOV} + t_0 - \frac{t_0}{P_s}\right) - \frac{t_0}{2} + t_0\ n \frac{(2-P_s)}{2P_s} \\ &= \frac{(2-P_s)}{2P_s}(n-1)t_0 + \tau_{FOV} \end{aligned} \tag{15A.8d}$$

where the second line is a mathematical simplification of the first line. This completes the derivation of (15.25).

Appendix 15B: Detection Times and Probabilities with Shared Information

In this appendix, (15.29), (15.32), (15.33), and (15.35) are derived.

Useful Mathematical Result

Let n_d denote the number of observers who detect a single stationary target in FOV search given as much time as needed. Denote the times these observers detect the target by $\mathbb{T}_1, \mathbb{T}_2,...,\mathbb{T}_{n_d}$. The $\mathbb{T}_i$ are random variables and, assuming all the observers are identical, they are drawn from a pdf $f(t)$. Define $\mathbb{T}_{\min}$ by the relationship

$$\mathbb{T}_{\min} \equiv min(\mathbb{T}_1, \mathbb{T}_2, \cdots, \mathbb{T}_{n_d}) \tag{15B.1}$$

where the function *min* picks out the argument with the smallest value. We seek the pdf for $\mathbb{T}_{\min}$. The probability that one of the $\mathbb{T}_i$ is between t and $t + dt$ and all the other $\mathbb{T}_i$ are greater than t is given by

$$f(t)dt\left(\int_t^\infty f(t')dt'\right)^{n_d-1}$$

Since there are n_d ways of choosing the variable $\mathbb{T}_i$ that is a minimum, the pdf for $\mathbb{T}_{\min}$ is

$$\begin{aligned} f_{\mathbb{T}_{\min}}(t) &= n_d f(t)dt\left(\int_t^\infty f(t')dt'\right)^{n_d-1} \\ &= n_d f(t)(1-F(t))^{n_d-1} \end{aligned} \tag{15B.2}$$

where $F(t)$ is the CDF associated with $f(t)$. Equation (15B.2) is given by Hsu[28] (p. 145). We now apply this result to find the pdf for $\mathbb{T}_{\min}$ assuming that $f(t)$ is described by (15.4a), which implies that $F(t)$ is given by (15.3a). Then (15B.2) becomes

$$\begin{aligned} f_{\mathbb{T}_{\min}}(t) &= n_d \frac{1}{\tau} e^{-\frac{t}{\tau}}\left(1-\left(1-e^{-\frac{t}{\tau}}\right)\right)^{n_d-1} \\ &= n_d \frac{1}{\tau} e^{-n_d\frac{t}{\tau}} \\ &= \frac{1}{\tau/n_d} e^{-\frac{t}{\tau/n_d}}, \quad 0 \le t < \infty \end{aligned} \tag{15B.3}$$

Amazingly, we find that the pdf for $\mathbb{T}_{\min}$ is exponential. Equation (15B.3) is a generalization of a result given by Papoulis.[42] From (15B.3) we readily find that

$$< \mathbb{T}_{\min} > = \frac{\tau}{n_d} \tag{15B.4}$$

The results of Section B.1 can be stated as a theorem: *Let $\mathbb{T}_i$ with $i = 1,\ldots,$ n_d be independent identical random variables all drawn from an exponential pdf with parameter τ and let $\mathbb{T}_{\min}$ be defined by (15B.1). Then the pdf of $\mathbb{T}_{\min}$ is given by (15B.3). From this result it immediately follows that $< \mathbb{T}_{\min} >$ is given by (15B.4).*

Mean Time for First Observer to Detect Target Given *n* Observers and P_∞

In this subsection we calculate the mean time to detect the target given n observers, τ, and P_∞. Our approach to deriving (15.29) is to derive an expression for $< \mathbb{T}_{\min} >$ when $n = 3$ and then generalize the result to calculate $< \mathbb{T}_{\min} >$ when n is arbitrary.

Suppose there are three observers. It is instructive to calculate the probabilities that zero, one, two or three observers detect the target, assuming identical independent observers:

- The probability that a single observer does not find the target is $1 - P_\infty$. This implies that the probability all three observers do not find the target is $(1 - P_\infty)^3$.
- The probability that the first observer finds the target but the second and third observers do not find the target is $P_\infty (1 - P_\infty)^2$. This can happen in three different ways so the probability that one observer finds the target but the other two do not is $3P_\infty (1 - P_\infty)^2$.
- The probability that the first two observers detect the target but the third observer does not is $P_\infty^2(1 - P_\infty)$. This can happen in three different ways so the probability that two observers detect the target but the third does not is $3P_\infty^2(1 - P_\infty)$.
- The probability that all three observers detect the target is P_∞^3.

The results are summarized below. Here $P(n)$ is the probability that n observers detect the target.

n	$P(n)$
0	$(1 - P_\infty)^3$
1	$3P_\infty(1 - P_\infty)^2$
2	$3P_\infty^2(1 - P_\infty)$
3	P_∞^3

One can readily check that these probabilities sum to one. Observe that the probability of one, two, or three observers detecting the target is $1 - (1 - P_\infty)^3$. Let

p_1, p_2, p_3 denote the fraction of the time one, two, or three observers detect the target, respectively. Then

$$p_1 = \frac{3P_\infty(1-P_\infty)^2}{1-(1-P_\infty)^3},\ \ p_2 = \frac{3P_\infty^2(1-P_\infty)}{1-(1-P_\infty)^3},\ \ p_3 = \frac{P_\infty^3}{1-(1-P_\infty)^3} \tag{15B.5}$$

Let $< \mathbb{T}_{\min}(n) >$ denote the mean time to detect the target when there are n observers. Then using (15B.4) and (15B.5), we find

$$\begin{aligned} < \mathbb{T}_{\min}(3) > &= p_1\tau + p_2\frac{\tau}{2} + p_3\frac{\tau}{3} \\ &= \tau\frac{3P_\infty(1-P_\infty)^2 + \frac{3}{2}P_\infty^2(1-P_\infty) + \frac{1}{3}P_\infty^3}{1-(1-P_\infty)^3} \end{aligned} \tag{15B.6}$$

The generalization of this result to n observers yields (15.29).

Mean Time to Detect Target with Two Observers Using Two Sensors

In this subsection we derive (15.33), which describes the case of two independent observers searching for a stationary target with two sensors. There are four cases:

1. Observer 1 and observer 2 find the target.
2. Observer 1 finds the target and observer 2 fails to find the target.
3. Observer 1 fails to find the target and observer 2 finds the target.
4. Observer 1 fails to find the target and observer 2 fails to find the target.

Let $\mathbb{T}_1$ and $\mathbb{T}_2$ denote the time the target is acquired by sensors 1 and 2. From (15.4a), $\mathbb{T}_1$ and $\mathbb{T}_2$ are random variables taken from an exponential distribution. Thus

$$< \mathbb{T}_1 > = \tau_1 \text{ and } < \mathbb{T}_2 > = \tau_2 \tag{15B.7}$$

Let $\mathbb{T}_{\min}$ denote the acquisition time. Then

- For case 1, $\mathbb{T}_{\min} = min\ (\mathbb{T}_1, \mathbb{T}_2)$.
- For case 2, $\mathbb{T}_{\min} = \mathbb{T}_1$.
- For case 3, $\mathbb{T}_{\min} = \mathbb{T}_2$.

The function $min\ (\mathbb{T}_1, \mathbb{T}_2)$ equals the smaller of $\mathbb{T}_1$ or $\mathbb{T}_2$.

Assume the detection probabilities for observers 1 and 2 are independent. Then

- The probability of case 1 is $P_{\infty 1}P_{\infty 2}$.
- The probability of case 2 is $P_{\infty 1}(1 - P_{\infty 2})$.

- The probability of case 3 is $P_{\infty 2}(1 - P_{\infty 1})$.
- The probability of case 4 is $(1 - P_{\infty 1})(1 - P_{\infty 2})$.

The reader can readily check that the sum of these four probabilities is one. It will subsequently be shown that

$$< min(\mathbb{T}_1, \mathbb{T}_2) > = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2} \tag{15B.8}$$

The probability for detection P_{det} is the sum of the probabilities for cases 1, 2, and 3:

$$\begin{aligned} P_{det} &= P_{\infty 1}P_{\infty 2} + P_{\infty 1}(1 - P_{\infty 2}) + P_{\infty 2}(1 - P_{\infty 1}) \\ &= P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2} \end{aligned}$$

- The probability for case 1, given detection, is $\dfrac{P_{\infty 1}P_{\infty 2}}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}}$.
- The probability for case 2, given detection, is $\dfrac{P_{\infty 1}(1 - P_{\infty 2})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}}$.
- The probability for case 3, given detection, is $\dfrac{P_{\infty 2}(1 - P_{\infty 1})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}}$.

Using (15B.7) and (15B.8) and the probabilities just computed, we find

$$< \mathbb{T}_{\min} > = \frac{P_{\infty 1}P_{\infty 2}}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}} \frac{\tau_1 \tau_2}{\tau_1 + \tau_2} + \frac{P_{\infty 1}(1 - P_{\infty 2})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}} \tau_1 + \frac{P_{\infty 2}(1 - P_{\infty 1})}{P_{\infty 1} + P_{\infty 2} - P_{\infty 1}P_{\infty 2}} \tau_2 \tag{15B.9}$$

which is the desired result.

It remains to show that (15B.8) is valid. Define a random variable $\mathbb{T}$:

$$\mathbb{T} = min(\mathbb{T}_1, \mathbb{T}_2) \tag{15B.10}$$

Then $\mathbb{T}$, $\mathbb{T}_1$ and $\mathbb{T}_2$ are random variables. We want to determine the pdf and expected value for $\mathbb{T}$. Equation (15B.3) applied to the case where the random variables were independent and drawn from an *identical* pdf. We cannot use (15B.3) here because the pdf for $\mathbb{T}_1$ is different from that for $\mathbb{T}_2$ since the first is characterized by τ_1 and the second is characterized by τ_2. Let $f_1(t;\tau_1)$ and $f_2(t;\tau_2)$ denote the pdf for observers using sensors 1 and 2 to detect the target. The probability that observer 1 detects the target between t and $t + dt$ and that observer 2 detects the target at a time later than t is

$$f_1(t)\,dt\left(\int_t^\infty f_2(t')\,dt'\right) = f_1(t)(1 - F_2(t))\,dt$$

Similarly, the probability that observer 2 detects the target between t and $t + dt$ and observer 1 detects the target at a time later than t is

$$f_2(t)(1 - F_1(t))\,dt$$

The pdf $f_{\mathbb{T}}(t)$ that either observer 1 or 2 detects the target at a time t and the other observer detects the target at a time greater than t is

$$f_{\mathbb{T}}(t) = f_1(t)(1 - F_2(t)) + f_2(t)(1 - F_1(t)) \tag{15B.11}$$

Apply (15B.11) to the problem at hand in which f_1 and f_2 are exponential with time constants τ_1 and τ_2:

$$\begin{aligned} f_1(t;\tau_1) &= \frac{1}{\tau_1} e^{-t/\tau_1} \quad F_1(t;\tau_1) = \left(1 - e^{-t/\tau_1}\right) \\ f_2(t;\tau_2) &= \frac{1}{\tau_2} e^{-t/\tau_2} \quad F_2(t;\tau_2) = \left(1 - e^{-t/\tau_2}\right) \end{aligned} \tag{15B.12}$$

Using these relationships, (15B.11) yields

$$\begin{aligned} f_{\mathbb{T}}(t) &= \frac{1}{\tau_1} e^{-t/\tau_1} \left(1 - \left(1 - e^{-t/\tau_2}\right)\right) + \frac{1}{\tau_2} e^{-t/\tau_2} \left(\left(1 - e^{-t/\tau_1}\right)\right) \\ &= \frac{1}{\tau_1} e^{-t\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)} + \frac{1}{\tau_2} e^{-t\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)} \\ &= \left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right) e^{-t\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)} \end{aligned} \tag{15B.13}$$

Define a τ appropriate to the pdf that describes $\mathbb{T}$ for the case where the pdf and CDF are given by (15B.12):

$$\frac{1}{\tau} \equiv \frac{1}{\tau_1} + \frac{1}{\tau_2} \quad \Rightarrow \tau = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2} \tag{15B.14}$$

Using this definition, (15B.13) is written in the form

$$f_{\mathbb{T}}(t) = \frac{1}{\tau} e^{-t/\tau} \quad 0 \le t < \infty \tag{15B.15}$$

Using (15B.10) and (15B.15), we get

$$< \mathbb{T} > = < min(\mathbb{T}_1, \mathbb{T}_2) > = \int_0^\infty t\, f_{\mathbb{T}}(t)\, dt = \tau$$

Using (15B.14), this shows that (15B.8) is valid, which in turn implies that (15B.9) is valid.

Probability Density Function for Detection Time with Two Observers Using Two Sensors

In this section we provide evidence that (15.35) is the appropriate pdf for the detection time when two observers search a FOV or FOR with two sensors. For convenience, rewrite (15.35) as follows:

$$f_{\mathbb{T}_{\min}} = \frac{P_{\infty 1}P_{\infty 2}}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}}\mathbb{E}\left(t;\frac{\tau_1\,\tau_2}{\tau_1+\tau_2}\right)+\frac{P_{\infty 1}(1-P_{\infty 2})}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}}$$

$$+\,\mathbb{E}(t,\tau_1)\frac{P_{\infty 2}(1-P_{\infty 1})}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}}\,\mathbb{E}(t,\tau_2) \tag{15B.16}$$

and recall that $\mathbb{E}(t;\tau)$ is the exponential pdf with parameter τ defined by (15.34).

We now show that in three limiting cases, (15B.16) yields the appropriate pdf:

- When $P_{\infty 1} = P_{\infty 2}$, the coefficient of the first term in (15B.16) is one and the remaining terms are zero. This corresponds to the case where the first and second observer eventually find the target. From (15B.13) the pdf for this case is $\mathbb{E}(t;\tau_1\tau_2/(\tau_1+\tau_2))$ in agreement with (15B.16).
- When $P_{\infty 1} = 1$ and $P_{\infty 2} = 0$, the coefficient of the second term in (15B.16) is one, which corresponds to sensor 1 detecting the target. From (15.4a) this corresponds to $\mathbb{E}(t;\tau_1)$ in agreement with the result from (15B.16).
- When $P_{\infty 1} = 0$ and $P_{\infty 2} = 1$, the coefficient of the third term in (15B.16) is one, which corresponds to sensor 2 detecting the target. From (15.4a) this corresponds to $\mathbb{E}(t;\tau_2)$ in agreement with (15B.16).

We have shown that the function defined by (15B.16) has the proper behavior in three limiting cases.

The exponential pdf defined by (15.34) has the property

$$\int_0^\infty \mathbb{E}(t;\tau)\,dt = 1 \;\text{ independent of }\; \tau \tag{15B.17}$$

If $f_{\mathbb{T}_{\min}}(t)$ is a pdf, then the integral over t must be 1.

$$\int_0^\infty f_{\mathbb{T}_{\min}}(t)\,dt = \mathbb{E}\frac{P_{\infty 1}P_{\infty 2}}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}}+\frac{P_{\infty 1}(1-P_{\infty 2})}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}}$$

$$+\frac{P_{\infty 2}(1-P_{\infty 1})}{P_{\infty 1}+P_{\infty 2}-P_{\infty 1}P_{\infty 2}} = 1 \tag{15B.18}$$

Because the area under $f_{\mathbb{T}_{\min}}(t)$ is one and $f_{\mathbb{T}_{\min}}(t)$ is always greater than or equal to zero, $f_{\mathbb{T}_{\min}}(t)$ is a legitimate pdf. The exponential pdf defined by (15.34) has the property

$$\int_0^\infty t\;\; \mathbb{E}(t,\tau)\,dt = \;\tau \tag{15B.19}$$

Multiplying (15B.16) by t, integrating over t, and using (15B.19) yields (15.33). The observation that $f_{\mathbb{T}_{\min}}(t)$ is a pdf, that it reduces to the correct pdf in three cases, and that it yields (15.33) strongly suggests that the pdf for $\mathbb{T}_{\min}$ is given by (15B.16).

Appendix 15C: *Mathematica* Search Code for TDSP Search Model

Define a function that is one between T_m and T_n and is zero elsewhere:

```
U[t_, Tm_, Tn_] := If[Tm < t ≤ Tn, 1, 0]
```

Define a function U_g that if plotted as a function of t, is a ramp function:

```
Ug[t_] := { t HeavisideTheta[t]   t ≠ 0
          { 0                     t == 0
```

Define a function W needed to describe the time dependence of $P(t)$ in the interval $T_m < t \le T_n$:

```
W[t_, T_, τ_] := (1 - Exp[-(t - T)/τ])
```

Define functions $P[i, t]$. Here i describes the interval and t describes the time dependence of $P[t]$ in the ith interval:

```
P[0, t_] = 0;

P[i_, t_] :=
 P[i, t] = (P[i - 1, T[i]] + Ug[Pω[i] - P[i - 1, T[i]]] W[t, T[i], τ[i]]) U[t, T[i], T[i + 1]]
```

Define $P_\infty[t, n]$:

```
Pω[t_, n_] := Sum[(Pω[i] U[t, T[i], T[i + 1]]), {i, 1, n}]
```

Define $\tau[t, n]$:

```
τ[t_, n_] := Sum[(τ[i] U[t, T[i], T[i + 1]]), {i, 1, n}]
```

Define the $P_r(t)$, $P_c(t)$, $p_r(t)$, and $p_c(t)$ functions:

```
Pr[t_, n_] := Sum[P[i, t], {i, 1, n}]

Pc[t_, n_] := Pr[t, n]/Pr[∞, n]

prDensity[t_, n_] := ∂t (Pr[t, n])

pcDensity[t_, n_] := ∂t (Pc[t, n])
```

Define a number of graphing routines:

```
P∞Plot[{P∞_, n_}, {t_, t0_, t1_}] := Module[{},
  Plot[P∞[t, n], {t, t0, t1}, AxesLabel → {"t", "P∞(t)"}, AxesOrigin → {0, 0},
   PlotRange → All, ImageSize → Medium, PlotStyle → {Red}]]

τPlot[{τ_, n_}, {t_, t0_, t1_}] := Module[{},
  Plot[τ[t, n], {t, t0, t1}, AxesLabel → {"t", "τ(t)"}, AxesOrigin → {0, 0},
   PlotRange → All, ImageSize → Small, PlotStyle → {Black}]]

PrPlot[{Pr_, n_}, {t_, t0_, t1_}] := Module[{},
  Plot[Evaluate[Pr[t, n]], {t, t0, t1}, AxesLabel → {"t", "Pr(t)"},
   AxesOrigin → {0, 0}, PlotRange → All, ImageSize → Small, PlotStyle → Blue]]

PcPlot[{Pc_, n_}, {t_, t0_, t1_}] := Module[{},
  Plot[Evaluate[Pc[t, n]], {t, t0, t1}, AxesLabel → {"t", "Pc(t)"},
   AxesOrigin → {0, 0}, PlotRange → All, ImageSize → Small, PlotStyle → Brown]]

prDensityPlot[{Pr_, n_}, {t_, t0_, t1_}] := Module[{PrDerivative},
  PrDerivative = ∂t (Pr[t, n]);
  Plot[PrDerivative, {t, t0, t1}, AxesLabel → {"t", "pr[t]"}, AxesOrigin → {0, 0},
   PlotRange → All, ImageSize → Small, PlotStyle → Blue]]

pcDensityPlot[{Pr_, n_}, {t_, t0_, t1_}] := Module[{PrDerivative},
  PrDerivative = ∂t (Pr[t, n]);
  Plot[PrDerivative / Pr[∞, n], {t, t0, t1}, AxesLabel → {"t", "pc[t]"}, AxesOrigin → {0, 0},
   PlotRange → All, ImageSize → Small, PlotStyle → Brown]]

DisplayAll[{Pr_, Pc_, P∞_, τ_, n_}, {t_, t0_, t1_}] :=
 Module[{g1, g2, g3, g4, g5, g6, g7, g8, g9},
  g1 = PrPlot[{Pr, n}, {t, t0, t1}];
  g2 = PcPlot[{Pc, n}, {t, t0, t1}];
  g9 = Show[{g1, g2}, AxesLabel → {"t", "Pr(t) & Pc(t)"}];
  g3 = P∞Plot[{P∞, n}, {t, t0, t1}];
  g4 = Show[{g1, g3}, AxesLabel → {"t", "P∞(t) & Pr(t)"}];
  g5 = prDensityPlot[{Pr, n}, {t, t0, t1}];
  g6 = pcDensityPlot[{Pr, n}, {t, t0, t1}];
  g7 = Show[{g5, g6}, AxesLabel → {"t", "pr(t) & pc(t)"}];
  g8 = τPlot[{τ, n}, {t, t0, t1}];
  GraphicsGrid[{{g4, g9}, {g7, g8}}]]
```

Write routines that take inputed TData, τData, and P∞Data and put it into arrays suitable for further processing:

```
LoadTData[TData_] := Module[{n, temp1},
  n = Length[TData];
  temp1 = Append[TData, ∞];
  Table[T[i] = temp1[[i]], {i, 1, n + 1}]]

LoadτData[τData_] := Module[{n},
  n = Length[τData];
  Table[τ[i] = τData[[i]], {i, 1, n}]]

LoadP∞Data[P∞Data_] := Module[{n, temp1},
  n = Length[P∞Data];
  temp1 = Prepend[P∞Data, 0];
  Table[P∞[i - 1] = temp1[[i]], {i, 1, n + 1}]]
```

References

[1] "The compendium of close combat tactical trainer algorithms, data, data structures and generic system mappings," *AMSAA Close Combat Analysis Division, Special Publication No. SP-97*, 2000.

[2] Blackwell, H., and D. McCready, "Foveal contrast thresholds for various durations of a single pulses," *Report Commander Bureau of Ships Contract No. Nobs-72038*, Vision Research Laboratories, Engineering Research Institute, University of Michigan, Ann Arbor, MI, 1946.

[3] D'Agostino, J., W. Lawson and D. Wilson, "Concepts for search and detection model improvements," *Proc. SPIE*, Vol. 3063, 1997, pp. 14–22.

[4] Deaver, D., S. Moyer, and C. Ra, "Modeling defined field of regard (FOR) search and detection in urban environments," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, G. Holst (ed.), *Proc. of SPIE*, Vol. 6207, 2006.

[5] Devitt, N., E. Flug, S. Moyer, B. Miller, and D. Wilson, "Search comparison in the infrared spectra for military operations in the urban terrain," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing* XV, G. Holst (ed.), *Proc. of SPIE*, Vol. 5407, 2004.

[6] Devitt, N., S. Moyer, and S. Young "Effect of Image Enhancement on the Search and Detection Task in the Urban Terrain," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, G. Holst (ed.), *Proc. of SPIE*, Vol. 6207, 2006.

[7] Doll, T., S. McWhorter, and A. Wasilewski, "Robust, sensor-independent target detection and recognition based on computational models of human vision," *Opt. Eng.*, Vol. 37, No. 7, 1988, pp. 2006–2021.

[8] Doll, T., S. McWhorter, D. Schmieder, and A. Wasilewski, "Simulation of selective attention and training effects in visual search and detection," in *Vision Models for Target Detection and Recognition*, E. Peli (ed.), River Edge, NJ: World Scientific Publishing Co., 1995.

[9] Doll, T., and R. Home, "Guidelines for developing and validating models of visual search and target acquisition," *Opt. Eng.*, Vol. 40, No. 9, 2001, pp. 1776–1783.

[10] Doll, T., R. Home, K. Cooke, A Wasilewski, D. Sheerin, et al., "A human vision simulation for evaluation of enhanced and synthetic vision systems," in *Enhanced and Synthetic Vision*, J. Verly (ed.), *Proc. SPIE*, Vol. 5081 2003, pp. 77–89.

[11] Driggers, R., P. Cox, and M. Kelley, "National imagery interpretation rating system and the probabilities of detection, recognition, and identification," *Opt. Eng.*, Vol. 36, No. 7, pp. 1952–1959, 1997.

[12] Driggers, R., P. Cox, J. Leachtenauer, R. Vollmerhausen, and D. Scribner, "Targeting and intelligence electro-optical recognition modeling: a juxtaposition of the probabilities of discrimination and the general image quality equation," *Opt. Eng.*, Vol. 37, No. 3, pp. 789–797, 1998.

[13] Driggers, R., K. Krapels, R. Vollmerhausen, P. Warren, D. Scribner, et al., "Target Detection Threshold in Noisy Color Imagery," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XII*, G. Holst (ed.), *Proc. of SPIE*, Vol. 4372, 2001.

[14] Driggers, R., E. Jacobs, R. Vollmerhausen, B. O'Kane, M. Self, et al., "Current infrared target acquisition approach for military sensor design and wargaming," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, G. Holst (ed.), *Proc. of SPIE*, Vol. 6207, 2006.

[15] Edwards, T., and R. Vollmerhausen, "Use of synthetic imagery in target detection model improvement," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XII*, G. Holst (ed.), *Proc. SPIE*, Vol. 4372, 2001.

[16] Edwards, T., R. Vollmerhausen, J. Cohen, and T. Harris, "Recent Improvements in Modeling Time Limited Search," in *Infrared and Passive Millimeter-wave Imaging Systems: Design, Analysis, Modeling, and Testing*, R. Appleby, G. Holst, and D. Wikner (eds.), *Proc. SPIE*, Vol. 4719, 2002.

[17] Edwards, T., R. Vollmerhausen, R. Driggers, and E. Grove, "NVESD Time-Limited Search Model," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XIV*, G. Holst (ed.), *Proc. SPIE*, Vol. 5076, 2003.

[18] Flug, E., T. Maurer, and O. Nguyen, "Time limited field of regard search," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, G. Holst (ed.), *Proc. SPIE,* Vol. 5784, 2005.

[19] Friedman, M., T. Du Bosq, J. Reynolds, D. Wilson, and R. Kang, "Comparison of perception results with a proposed model for detection of a stationary target from a moving platform," in *Infrared Imaging Systems: Design, Analysis, Modeling and Testing XX*, G. Holst (ed.), *Proc. SPIE*, Vol. 7300, 2009.

[20] Friedman, M., T. Du Bosq, and E. Flug, "Analytical Models Quantify Military Benefit of Collaborative Search," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI*, G. Holst and K. Krapels (eds.), *Proc. of SPIE*, Vol. 7662, 2010.

[21] Friedman, M., and J. Reynolds, "Collaborative search with independent sensors and multiple observers," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, G. Holst and K. Krapels (eds.), *Proc. of SPIE*, Vol. 8014, 2011.

[22] Friedman, M., J. Reynolds, D. Wilson, and R. Driggers, "Analytical calculation for probability of detection given time dependent search parameters," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, G. Holst and K. Krapels (eds.), *Proc. of SPIE*, Vol. 8014, 2011.

[23] Handford, M., *Where's Waldo?*, Boston, MA: Little Brown and Company, 1987.

[24] Handford, M., *The Great Waldo Search*, Little Brown and Company, 1989.

[25] Hogervorst, M., P. Bijl, and A. Toet, "On the relationship between human search strategies, conspicuity and search performance," in *Infrared Imaging Systems: Design, Analysis, Modeling and Testing XVI*, G. Holst (ed.), *Proc. SPIE*, Vol. 5784, 2005, pp. 240–251.

[26] Holst, G., *Electro-Optical Imaging System Performance, Fifth Edition*, Bellingham, WA: SPIE Press, 2008.

[27] Howe, J., "Electro-Optical imaging system performance prediction," in *Electro-Optical Systems Design, Analysis, and Testing*, M. Dudzik (ed.), Bellingham, WA: SPIE Press, 1993.

[28] Hsu, H., *Probability, Random Variables and Random Processes*, New York City, NY: McGraw Hill, Schaum Outline Series, 1997.

[29] Irvine, J., "National imagery interpretability rating scales (NIIRS): overview and methodology," in *Airborne Reconnaissance XXI*, W. Fishell (ed.), *Proc SPIE*, Vol. 3128, 1997, pp. 93–103.

[30] Koopman, B., *Search and Screening: General Principles and Historical Applications*, Tarrytown, NY: Pergamon Press, 1980. Republished in 1999 by the Military Operations Research Society, Inc., Alexandria, VA.

[31] Leachtenauer, J., and R. Driggers, *Surveillance and Reconnaissance Imaging Systems: Modeling and Performance Prediction*, Norwood, MA: Artech House, 2001.

[32] Maurer, T., R. Driggers, and D. Wilson, "Search and detection modeling of military imaging systems," *Infrared Imaging Systems: Design, Analysis, Modeling and Testing XVI*, G. Holst (ed.), *Proc. SPIE*, Vol. 5784, 2005, pp. 201–215.

[33] Maurer, T., D. Wilson, S. Smith, D. Deaver, E. Flug, et al., "Search and detection comparing midwave and longwave infrared," *Opt. Eng*., Vol. 48, No. 11, 2009.

[34] Meitzler, T. J., G. R. Gerhart, and S. Harpreet, "Advances in target acquisition modeling," *Opt. Eng.*, Vol. 37, 1998, pp. 1900–1901.

[35] Moore, R., H. Camp, S. Moyer, and C. Halford, "Triangle search experiment to isolate scene clutter effects," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI*, G. Holst and K.. Krapels (eds), *Proc. of SPIE*, Vol. 7662, 2010.

[36] Moore, R., H. Camp, S. Moyer, and C. Halford, "Masked target transform volume clutter metric applied to vehicle search," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXI*, G. Holst and K. Krapels (eds.), *Proc. of SPIE*, Vol. 7662, 2010.

[37] Nguyen, O., E. Flug, and T. Maurer, "The effect of targets in defilade on the search task," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, G. Holst (ed.), *Proc. of SPIE*, Vol. 5784, 2005.

[38] Nicoll, J., "A mathematical framework for an improved search model," *IDA Paper P-2901*, Institute for Defense Analysis, Alexandria, VA, 1994.

[39] Nicoll, J., J. Cartier, and D. Hsu, "Neoclassic model for Search," in *Electro-Optical Imaging: System Performance and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[40] O'Kane, B., and G. Page, "Angular distance traveled across the eye as figure of merit for detecting moving targets," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVII*, G. Holst (ed.), *Proc. of SPIE*, Vol. 6207, 2006, 620701.

[41] Papoulis, A., S. Pillai, *Probability, Random Variables and Stochastic Processes, Fourth Edition*, New York City, NY: McGraw Hill, 2002.

[42] Peebles, P., *Probability*, *Random Variables and Random Signal Principles*, New York City, NY: McGraw Hill, 1993.

[43] Rotman, S., E. S. Gordon, and M. Kowalczyk, "Modeling human search and target acquisition performance: 1. First detection probability in a realistic multitarget scenario," *Opt. Eng.*, Vol. 28, No. 11, 1989, pp. 1216–1222.

[44] Rotman, S., "Modeling human search and target acquisition performance: II. Simulating multiple observers in dynamic scenarios," *Opt. Eng.*, Vol. 28, No. 11, 1986, pp. 1223–1226.

[45] Rotman, S., "Modeling human search and target acquisition performance: III. Target detection in the presence of obscurants," *Opt. Eng.*, Vol. 30, No. 6, 1991, pp. 824–829.

[46] Rotman, S., "Modeling human search and target acquisition performance: IV. Detection probability in the cluttered environment," *Opt. Eng.*, Vol. 33, No. 3,1994, pp. 801–808.

[47] *Search Theory: Some Recent Developments*, D. Chudnovsky and G. Chudnovsky (eds.), CRC Press, 1988.

[48] Silk, J., "Modeling the observer in target acquisition," *IDA Paper P-3102*, Institute for Defense Analysis, Alexandria, VA, 1997.

[49] Silk, J., "Modeling the observer in target acquisition," in *Electro-Optical Imaging: System Performance and Modeling*, L. Biberman (ed.), Andover, MA: ONTAR Corp., 2000.

[50] Stone, L., *Theory of Optimal Search*, New York City, NY: Academic Press, 1975.

[51] Vollmerhausen, R., R..Driggers, and B. O' Kane, "Influence of sampling on target recognition and identification," *Opt. Eng.*, Vol. 38, No. 5, pp. 763–772, 1999.

[52] Vollmerhausen, R., J. Olson, and J. Reynolds, "Modeling the benefit of color in target acquisition: characterizing color vision," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XIX*, G. Holst (ed.), *Proc. of SPIE*, Vol. 6941, 2008.

[53] Wilson, D., N. Devitt, and T. Maurer, "Search times and probability of detection in time-limited search," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, G. Holst (ed.), *Proc. of SPIE*, Vol. 5784, 2005.

CHAPTER 16

Laboratory Measurements of Infrared Imaging System Performance

To this point we have discussed the modeling of various aspects of infrared and EO systems. As in any other scientific discipline, it is essential for system performance to be characterized and the utility of the model assessed; that is, how well does it predict observed performance? For example, an MWIR camera is partially characterized by NETD, MTF, and MRTD (or MRT). These are all quantities that can be estimated from laboratory data and compared to those predicted by the system model. Models with good predictive ability can be subsequently used to forecast imager performance (e.g., probability of detection and identification as a function of range) under a variety of useful scenarios.

In the following sections, we review the basic IR system performance parameters and describe their estimation from laboratory measurement. An accurate assessment of both spatial and temporal noise levels is essential because these quantities place fundamental physical limits on the sensitivity and resolution of an imager. The issues associated with human performance testing are also discussed and we consider both the standard MRT calculation and the more recently developed dynamic MRT (DMRT). Work and future trends for sampled imaging system measurements are also briefly discussed.

16.1 Sensitivity

The classical infrared performance measure that describes the sensitivity of a sensor is called the *noise equivalent temperature difference* (NETD) [1, 2, 3]. As the name implies, the NEDT is the amplitude of a step change in temperature (difference from background) that results in a SNR of unity. Thus, the estimation of NEDT requires both "signal" and "noise" components. Here we discuss the estimation of NEDT for a staring MWIR camera. For a staring array the SNR can be defined for each pixel i as

$$SNR_i(T) = \left|\Delta Q_i(T)\right| / \sigma_i(T) \tag{16.1}$$

where $\Delta Q_i(T) = Q_i(T + \Delta T) - Q_i(T)$ is the difference in the incident number of photons caused by a signal of ΔT K. The source of uncertainty in this case is temporal noise, characterized here by the pixel noise standard deviation $\sigma_i(T)$. Assuming the step change in temperature is small, we may expand the signal in a Taylor series:

$$Q_i(T + \Delta T) = Q_i(T) + \Delta T\, dQ_i(T)/dT + \ldots \tag{16.2}$$

so that to leading order, $\Delta Q_i(T) = \Delta T\ dQ_i(T)/dT$. Substituting into (16.1) we can solve $SNR_i(T) = 1$ for ΔT to yield (in kelvin)

$$NEDT_i(T) = \frac{\sigma_i(T)}{\frac{dQ_i(T)}{dT}} \tag{16.3}$$

Thus, the pixel-level NETD depends on both the temporal system noise and on the instantaneous rate of change in incident radiation with respect to background temperature. By definition, the NETD is therefore not a fixed quantity but will depend on the ambient temperature of the scene, T. Although the dependence has not been made explicit, the NEDT also will depend on other camera parameters, most notably the frame integration time and transmission of the optics. Longer integration times, for example, tend to increase the SNR. Here we have assumed the integration time has been fixed to a desired value.

Estimation of the NEDT therefore requires measuring both the noise and a known signal as a function of temperature. With regard to the latter, a calibrated blackbody source is the instrument of choice and is shown in Figure 16.1.

The source is set to a fixed temperature and the imager under test is oriented such that the square target occupies a large portion of the camera field of view. A sequence of M frames (images) is collected so that estimates of the temporal variance can be obtained for each pixel using

$$\hat{\sigma}_i^2 = \frac{1}{M-1}\sum_{m=1}^{M}(Q_{im} - \bar{Q}_i)^2 \tag{16.4}$$

where Q_{im} is the number of photons recorded for pixel i and frame m. The overbar denotes the average number of photons recorded at pixel i over the same M frames. This process is repeated for different blackbody temperatures T_k for $k = 1\ldots K$ yielding the estimates $\sigma_i(T_k)$ and $Q_i(T_k)$. As a final step, both quantities can be fit to a low-order polynomial in T. This provides a good way to estimate the derivative dQ_i/dT and also allows us to predict the NEDT at temperatures other than those

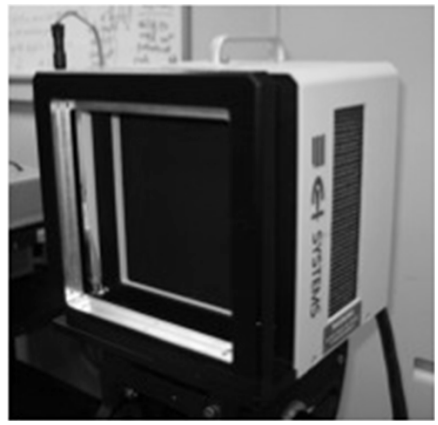


Figure 16.1 Blackbody source used to generate known (calibrated) signals for IR camera system testing.

at which data were recorded. More specifically, we fit $\sigma_i(T_k) = b_0 + b_1 T_k + b_2 T_k^2$ and $Q_i(T_k) = a_0 + a_1 T_k + a_2 T_k^2$ for the coefficients a and b from which we arrive at the following estimate:

$$NEDT_i(T) = \frac{b_0 + b_1 T + b_2 T^2}{a_1 + a_2 T} \tag{16.5}$$

As an example, consider the panoramic MWIR imager shown in Figure 16.2. At the core of this imager is a 2048×2048 InSb detector array. The fore-optics comprises several mirrors and lenses (i.e., a catadioptric system) which serve to properly focus a 360-deg azimuth scene onto the FPA. Other details of the imager can be found in [2].

The blackbody target was set to a series of 10 temperatures ranging from 270K to 315K. At each temperature the imager recorded a sequence of M = 50 frames with the target largely covering the entire FOV. From these data, the above-described estimation procedure was used to produce NETD values for each pixel. Figure 16.3 shows the distribution of these estimated NETD values at room temperature (293K). The vertical line shows the model-predicted NETD, obtained using the methods described in earlier chapters. The model assumed background limited performance such that the temporal noise standard deviation is modeled as the square root of the signal strength [2].

Note that these calculations can be performed both with and without the system optics present. Because the camera optics filter the incoming radiation, it is useful to perform the test both with optics present and by directly "flood illuminating" the FPA with no intervening optics. The differences in NETD values can be significant (see, e.g., [2]).

Note also that nonuniformities of the detector arrays can contribute significantly to the overall noise of the system [see, e.g., Figure 16.3(a)]. In fact, this "spatial noise" gave rise to the spatial distribution of the NETD values seen in Figure 16.3(b). An interesting question becomes what single NETD value should one use for a staring system? The answer requires a thorough accounting of the various sources of uncertainty present in today's imaging systems. For the calculations above we took the noise variance to be that associated with temporal variations only. However, the

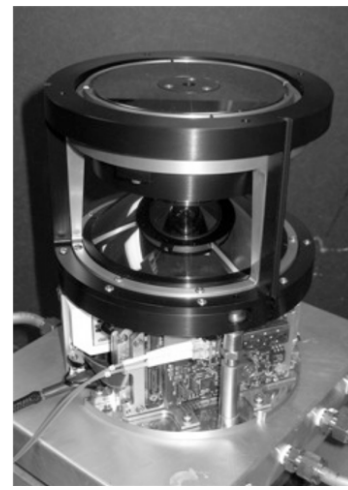


Figure 16.2 Panoramic MWIR imager used in the example calculations performed in this chapter.

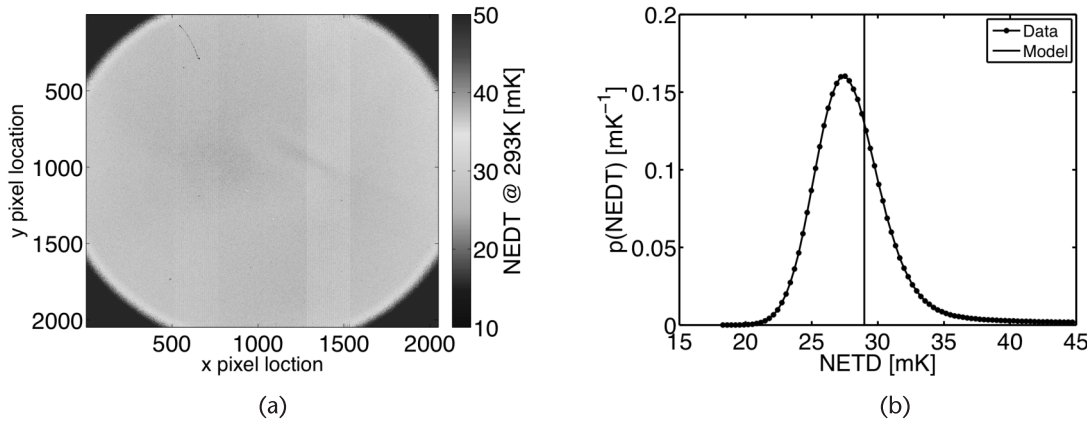


Figure 16.3 (a) Flood-illuminated portion of the FPA and the pixel-level estimates of the NETD. There is clearly spatial variability in the noise performance of the FPA. (b) Probability density (spatial) of NETD values. The mean of the distribution is close to the value predicted by the camera model, obtained using the methods presented in earlier chapters.

NETD should more properly reflect any source of uncertainty that would prevent our ability to detect a small change in temperature, including spatial noise.

Figure 16.4 shows the spatial noise standard deviation (normalized by signal strength) for our camera system. These values were estimated after applying a two-point NUC. Folding spatial variance into an NETD calculation is straightforward, pending certain assumptions on the nature of the noise. If we assume that both spatial and temporal noise are approximately Gaussian distributed, we may replace the pixel-level noise term in (16.3) with the total spatiotemporal noise variance $\sigma_{Tot}^2 = \sigma_t^2 + \sigma_s^2$. In this expression we take the temporal noise variance σ_t^2 as the average temporal variance over all pixels. The Gaussian assumption on the noise

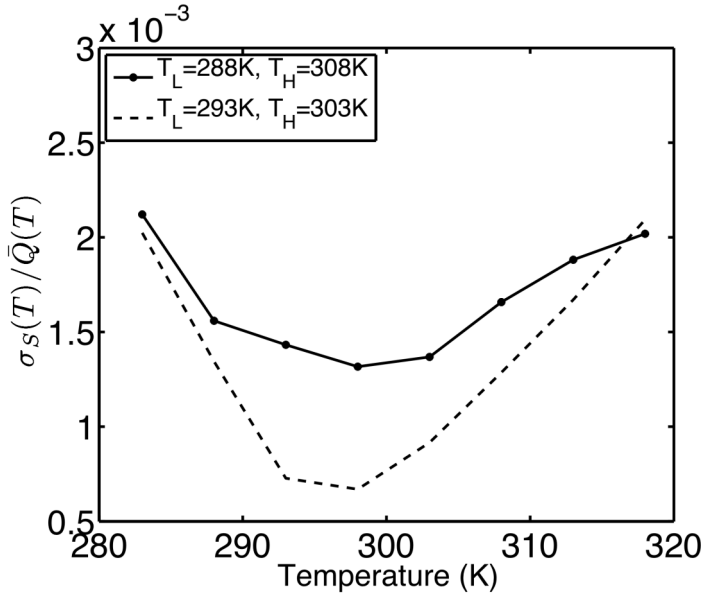


Figure 16.4 Normalized spatial noise σ_s as a function of temperature for post-NUC imagery. The different curves correspond to different calibration temperatures (high and low) used in the NUC. Spatial noise contributes a small amount to the overall noise level for this system.

is what allows us to add variances (other noise models need to be combined differently). Replacing the pixel-level variance in 16.3 with σ_{Tot} and the signal term dQ_i/dT with the spatially averaged signal derivative provides a single NETD measure for the camera. Combining both spatial and temporal uncertainty provides a more realistic NETD value that is more in line with the intent of the measurement.

Note also that computing NETD *prior* to NUC would result in much higher NETD values. This can be seen by inspection of Figure 16.3(a) where the spatial nonuniformities are clearly visible. If spatial nonuniformities or *fixed-pattern noise* is accounted for in the numerator of (16.3) the result is sometimes referred to as the inhomogeneity equivalent temperature difference (IETD) [3].

Table 16.1 describes several types of uncertainty and their source. Each of these noise sources will have an associated standard deviation that can be combined in the same way as was done for σ^2_{Tot} above (sum of component variances). A more thorough accounting of noise sources in laboratory measurements including spatiotemporal, or 3D, noise can be found in [3].

Finally, we should also mention that the process for estimating NETD may vary depending on the type of imager used. Scanning systems, for example, may require a slightly different approach (see, e.g., [3]).

16.2 Resolution

The MTF is a measure of system resolution, or the resolution of one of the system components[4,5]. Theoretically, MTF is the system magnitude response to a set of input sinusoids of various spatial frequencies. Ideally we would present the camera with a series of spatially varying, sinusoidal patterns and record the amplitude of the measured response. However, this is quite difficult to do experimentally due to noise in the measured data; hence, this so-called "direct" approach is seldom used. Methods based on Fourier analysis are far more common.

Recall that for a linear shift-invariant system, the image of a point source target is the impulse response, or the point spread function, *psf*(*x*,*y*), of the system [6]. As discussed in Chapter 3, we could measure this quantity and take the 2D Fourier transform to obtain the MTF. The issue with this approach is the degree to which one can produce a signal approximating a true point source. A better option is to measure the l-D MTF by using an edge target. This procedure is shown schematically in Figure 16.5 for a scanning imager. The response to a line target is called the line spread function, *lsf*, and the Fourier transform of the *lsf* is equal to the l-D MTF. Note that this is *not* equivalent to a "slice" out of the 2D MTF. The

Table 16.1 Sources of Spatiotemporal Noise

Component	*Potential Source*
Random spatiotemporal noise	Detector temporal noise
Temporal row bounce	Line processing, 1/*f*, readout
Temporal column bounce	Scan effects
Random spatial noise	Pixel processing, detector nonuniformity
Fixed row noise	Detector nonuniformity, 1/*f*
Fixed column noise	Detector nonuniformity, scan effects
Frame to frame noise	Frame processing

use of an edge target can help in at least two ways: (1) More signal response is obtained and (2) target size is not an issue as it is for a point source.

Define the edge spread function $esf(x,y)$ as the response of the imaging system to the edge target. The *esf* is related to the *lsf* in the direction normal to the edge by

$$lsf(x) = \frac{d[esf(x)]}{dx} \tag{16.6}$$

Unlike the *psf*, the *esf* must be measured in both the horizontal and vertical directions to obtain the horizontal and vertical MTF curves. Note that separability is assumed for the edge spread function (see Chapter 2 for the assumption of separability).

Using the ESF to estimate the MTF requires an edge model. The line spread function can often be well approximated by a Gaussian distribution in each spatial direction. Integrating therefore provides a model for the ESF given as

$$esf(x) = \frac{1}{2} + \frac{1}{2} erf[x/\sigma] \tag{16.7}$$

where the variance of this distribution depends on the system *f*-number and wavelengths being imaged. Thus, to identify the ESF we record an edge response as in Figure 16.5, normalize the amplitude of the response to unity, and fit the data to (16.7) to obtain the standard deviation σ. This effectively identifies the $lsf(x)$ in one

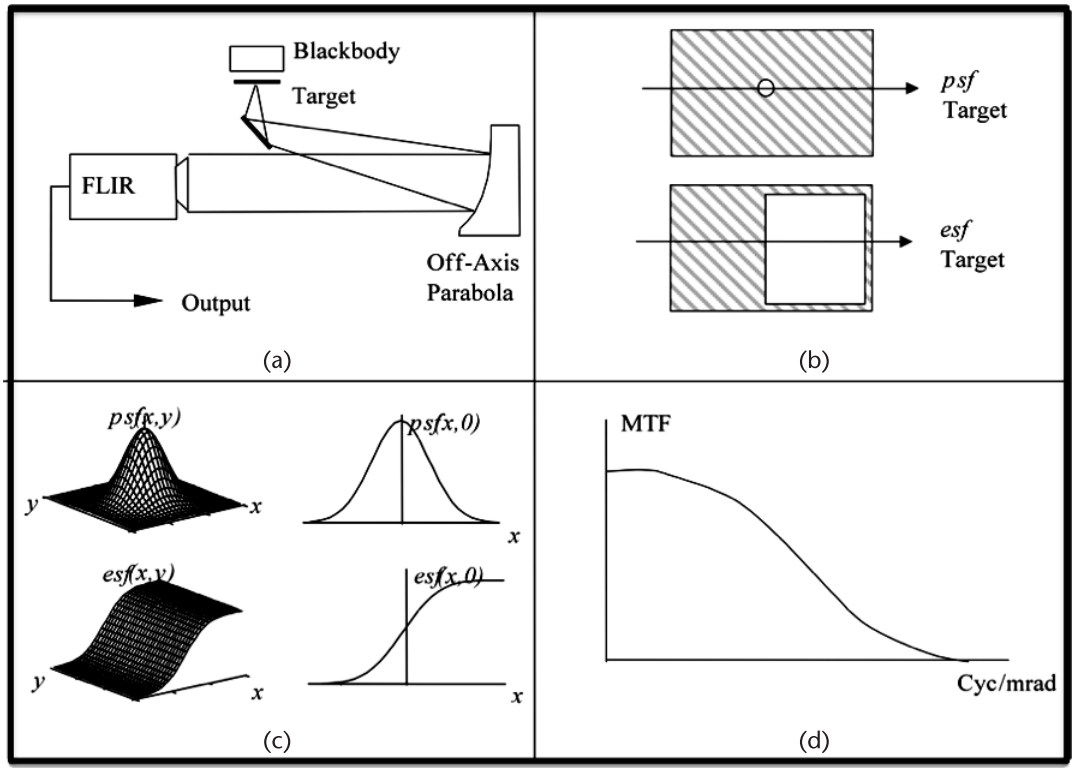


Figure 16.5 MTF measurement process.

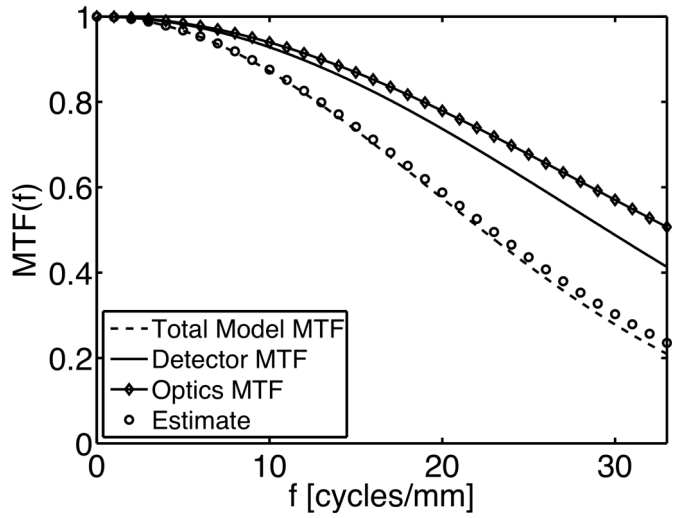


Figure 16.6 Measured MTF for the MWIR camera system of Figure 16.2. The measured MTF can be accurately modeled by including both optics and detector MTF components only.

dimension via (16.6). Taking the Fourier transform therefore provides the desired horizontal MTF. The process can be repeated in the vertical direction to obtain the associated MTF. Usually, the spatial axis of the *psf* (or *lsf*) is in milliradians and the Fourier transform gives a frequency domain function with units of cycles per milliradian.

The procedure is the essentially the same for a staring imager. Rather than scanning across an edge target, one simply places an edge in the FOV and records the edge response directly. The total MTF for the MWIR camera of Figure 16.2 was estimated using this procedure and is shown by the open circles in Figure 16.6. Recall, however, that the experimentally obtained MTF is modeled as a linear combination of several components including optics, electronics, and detector (see Chapter 11). Therefore, the measured MTF is described by the product

$$MTF(\xi,\eta) = H_{optics}(\xi,\eta)H_{detector}(\xi,\eta)H_{electronics}(\xi,\eta)H_{display}(\xi,\eta)\ldots \tag{16.8}$$

For example, Figure 16.6 shows both the optics and detector transfer functions as well as the total MTF. In this case the product of these two transfer functions accurately describes the MTF. However, in this analysis the measured MTF did not include the influence of the system display. This is important since the display can be the limiting component of the system. The eye impulse response, while included in performance models, was not included in the overall system MTF measurement. This brings us naturally to the MRTD, where the human eye–brain response *is* included in the measurement.

16.3 Human Performance: Minimum Resolvable Temperature Difference

The ultimate measure of thermal imaging system performance is a measure of the visual acuity of a human observer through the sensor under test. The measurement is called the *minimum resolvable temperature difference* (MRTD or just MRT). [7–9] MRT is a

measure of an observer's ability to resolve a four-bar pattern through a sensor under test. Because results depend on observers' training, motivation, and visual acuity, several trained observers are required [10].

Figure 16.7(a) shows the test configuration with an observer viewing the target image on a display. For characterizing a particular system, the most accurate results are obtained if the display is the actual display to be used with the system in the field, but for comparing systems it may be advantageous to use the same high-quality display monitor. A 7:1 aspect ratio, four-bar target is positioned with the bars oriented in the vertical direction in order to measure the horizontal MRT. The azimuth and elevation of the system under test are generally adjusted to center the target in the field of view. The observer is allowed to adjust the monitor and system level and gain during the test. A negative contrast (the bars are cooler than the background) sub-MRT temperature differential is obtained first. The differential temperature is slowly decreased until the observer can resolve all four bars. Then, the temperature difference is recorded at which the observer can resolve all four bars in positive contrast.

The absolute value of the two measurements is averaged and taken as the MRT for the observer at the spatial frequency of the target. Note that the ambient temperature/emissivity offset is the value in between those two temperature values. The same procedure is performed for all spatial frequency targets. Usually, there are five to eight different targets in the set. Each observer completes up to three runs (if the first two are not consistent) and averages the MRTD values for each of the spatial frequencies in the target set. Then, the MRTD values for all of the observers are

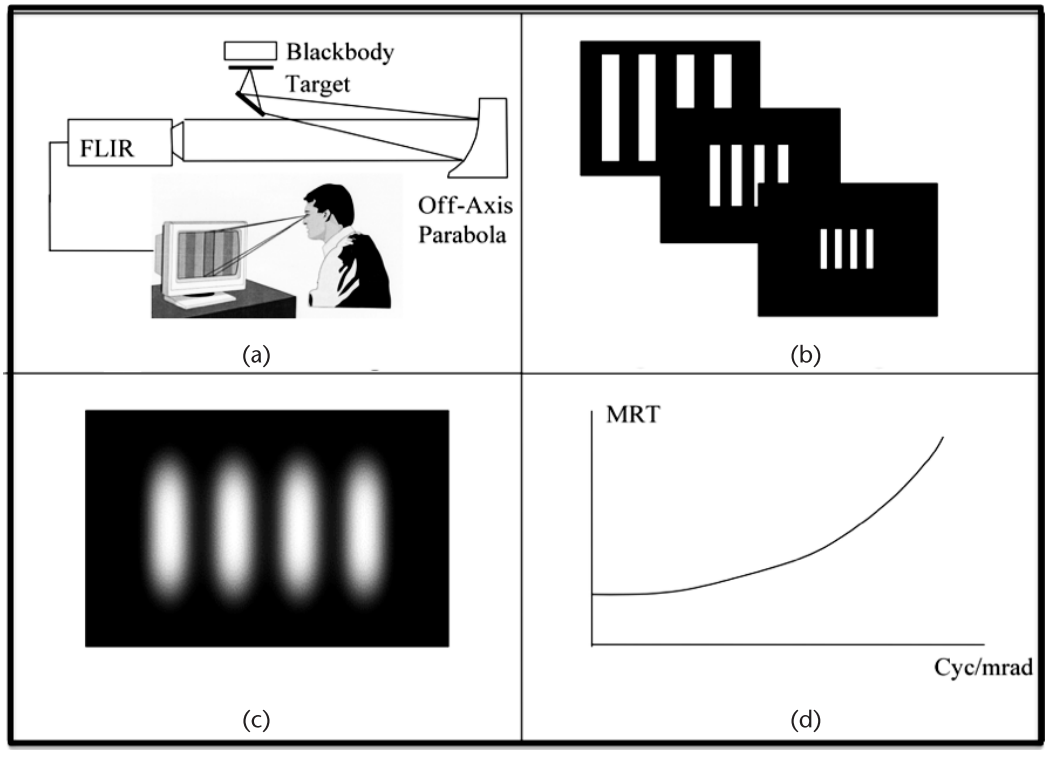


Figure 16.7 MRT measurement process.

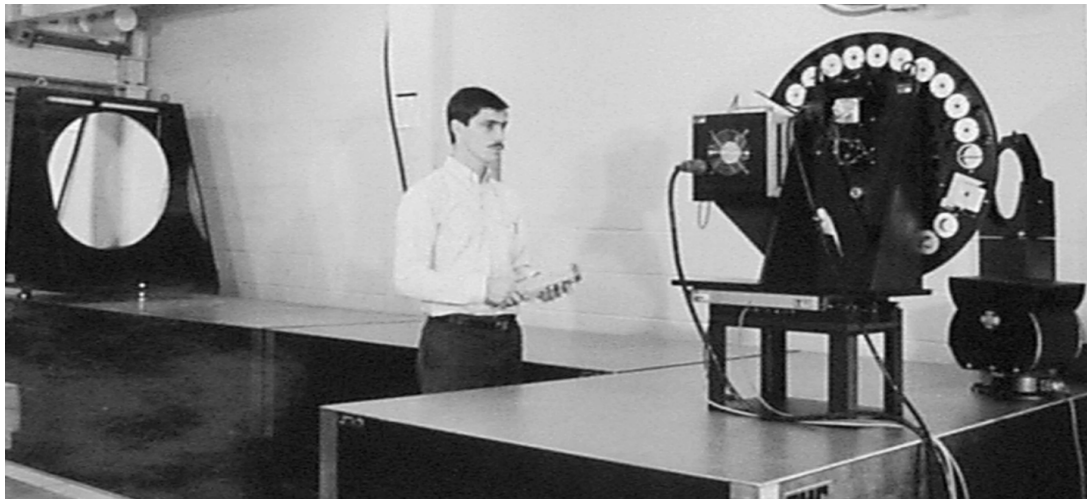


Figure 16.8 Target wheel at focal plane of collimator.

averaged resulting in one curve. This is repeated for both horizontal and vertical bar orientations. The 2D MRTD curve is calculated from a geometrical average of spatial frequencies corresponding to particular temperature values from the horizontal and vertical MRT curves. The final measurement is shown in Figure 16.7(d).

Figure 16.8 shows a target wheel with a large number of target selections. This particular wheel is so large and heavy that a crane is needed to move the wheel into position. The target wheel is positioned at the focal plane of an off-axis paraboloid as shown in the figure. A fold mirror is used to position the target wheel at the focal plane.

16.4 Dynamic Minimum Resolvable Temperature Difference

The characteristics of sampled imaging systems have been a serious issue with MRT measurements [11, 12]. With traditional MRT, a phenomenon occurs between 0.6 and 0.9 times the half-sample (Nyquist) frequency of the sensor. In this region, the transfer response for sampled imaging systems can be degraded significantly or even enhanced. Also, depending on the subjective criteria adopted at a given test facility, the MRT can be measured beyond the half-sample rate of the sensor under test.

Figure 16.9 shows the static approach to measuring the MRT of a staring sensor. At each frequency or bar pattern, the phase is found that optimizes the MRT. To vary the phase, the target wheel is placed on a translating stage or, equivalently, the sensor is placed on a motorized rotational stage. The phase is considered the angular target position with respect to the sensor detector. While the observer views the four-bar target, the stage is adjusted (i.e., the phase is varied) or the sensor is micropositioned to pick up the four-bar target modulation. In the region between 0.6 and 0.9 times the half-sample rate of the sensor, it is extremely difficult to ensure that the correct target phase is accomplished. The results vary greatly with phase and from observer to observer. Figure 16.9(b) shows two images of a four-bar target where one target is positioned at the best phase and the other

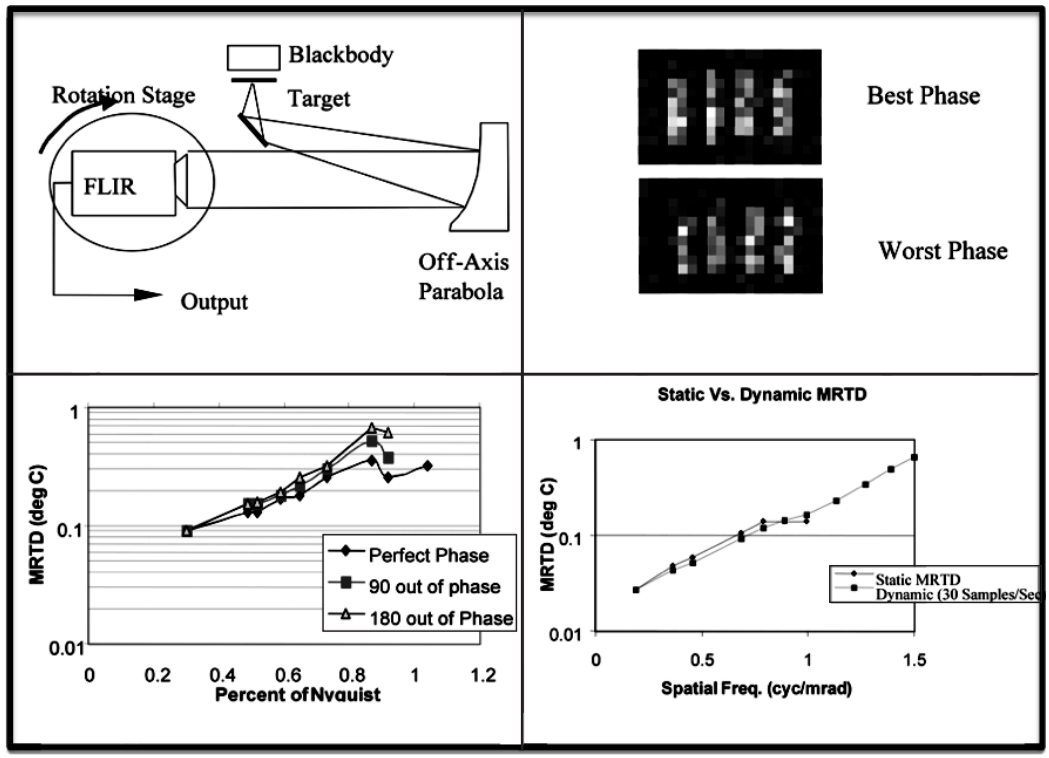


Figure 16.9 Sample imaging system MRT measurements.

target is positioned at the worst phase. Figure 16.9(c) shows the MRT for a sensor at best phase and at worst phase. Also note in the static MRT figure, the MRT is measured past the half-sample rate of the sensor for the best phase case [13].

To address the phase and sampling problems associated with the MRT measurement, Webb [14] developed the dynamic MRT measurement where the MRT target is moved across the sensor field of view during the MRT measurement. The technique is known as the *dynamic MRT* (DMRT) and the MRT is measured as the target is moved through the sensor's FOV.

The DMRT has been shown capable of measuring system performance up to the MTF limit of the system and eliminates the problems associated with static sampling and phase optimization. The increase in MRT performance is measured due to an increase in the sample rate across the target (as the target moves across the sampling grid).

Two characteristics are worth mentioning here. First, while the sampling effect has been minimized in the DMRT measurement, additional MTF blur has been introduced by the motion of the target. The DMRT measurement changes with target velocity through the sensor FOV since both the spatial sampling rate and the blur due to motion have changed. The DMRT measurement beyond the half-sample rate of the sensor is useful and can be modified for use in acquisition calculations. Webb and Halford describe the DMRT method in detail [15].

A great deal of work has yet to be performed in the laboratory measurement of sampled imaging systems. The international community is working on many issues

associated with the performance measurement of sampled imaging systems. Until these issues are addressed, engineers and acquisition professionals should be cautious about laboratory results and claims of performance with sampled imaging systems. The interpretation of these measurements is taken many different ways and will continue to be taken different ways until a common approach is established.

References

[1] Lloyd, J. M., *Thermal Imaging Systems*, NY: Plenum Press, 1975, pp. 166–167.

[2] Nichols, J. M., J. R. Waterman, R. Menon, and J. Devitt, "Modeling and Analysis of a High-Performance Midwave Infrared Panoramic Periscope," *Optical Engineering*, Vol. 49, No. 11, 2010, 113202 .

[3] Driggers, R. G., C. Webb, S. J. Pruchnic Jr., C. E. Halford, and E. E. Burroughs, Jr., "Laboratory Measurement of Sampled Infrared imaging System Performance," *Optical Engineering*, Vol. 38, No. 5, 1999, pp. 852–861.

[4] Pinson, L., *Electro-Optics*, NY: Wiley, 1985, p. 119.

[5] Boreman, G., *Handbook of Optics*, NY: McGraw-Hill, 1995, p. 32.8.

[6] Gaskill, J., *Linear Systems, Fourier Transforms, and Optics*, NY: Wiley, 1978, pp. 343–345.

[7] Holst, G., *The Infrared and Electro-Optical Systems Handbook—Volume 4 Infrared Systems Testing*, SPIE and ERIM, 1993, p. 235.

[8] Hoover, C., and C. Webb, "What is an MRT and how do I get one?," *SPIE Proc.*, Vol. 1488, 1991, pp. 280–288.

[9] Webb, C., "MRTD, how far can we stretch it?," *SPIE Proc.*, Vol. 2224, 1994, pp. 294–304.

[10] Webb, C., and G. Holst, "Observer variables in MRTD," *SPIE Proc.*, Vol. 1689, 1992, pp. 356–367.

[11] Webb, C., "Results of laboratory evaluation of staring arrays," *SPIE Proc.*, Vol. 1309, 1990, pp. 271–278.

[12] Ferrari, G. B., *Measurement of the Minimum Resolvable Temperature Difference (MRTD) of Thermal Cameras*, Standard NATO Agreement 4349, 1994.

[13] Vollmerhausen, R., R. Driggers, C. Webb and T. Edwards, "Staring imager minimum resolvable temperature (MRT) measurements beyond the sensor half sample rate," *Opt. Eng.*, Vol. 37, No. 6, 1998, pp. 1763–1769.

[14] Webb, C., "Dynamic minimum resolvable temperature difference for staring focal plane arrays," *IRIS Passive Sensors Conference*, 1993.

[15] Webb, C., and C. Halford, "Dynamic minimum resolvable temperature difference for staring array imagers," *Opt. Eng.*, Vol. 38, No. 5, 1999.

List of Symbols

α	absorptance. See (5.1)
α	temperature coefficient of resistance. See (8.11)
β	a calibration constant approximately equal to $5.0\sqrt{Trolands\ \sec}$. See (13.5b)
β	angular size of diffraction blur. See Figures 7.70 and 7.71
γ	direction cosines. See Figure 4.3.
γ	a dimensionless constant approximately equal to 330. See (13.5c).
γ	normalized correlation function. See (4.14).
γ_g	autocorrelation of g. See Table 2.3.
γ_{fg}	correlation of functions f and g. See Table 2.3
γ	scattering coefficient. See (6.9)
ε	emittance. See (5.1), (5.15), (5.16)
$\in$	a dimensionless number used in dimensional analysis. See (13.23a).
η	spatial frequency in vertical direction. [cycles/length] or [cycles/mr]
λ	wavelength. See Figure 4.2.
μ	mean. See (2.65a and (5.47)
υ	temporal frequency
ξ	spatial frequency in the horizontal direction. [cycles/length] or [cycles/mr]. See Figure 13.2
π	a dimensionless variable. See (13.24a).
ρ	a spatial frequency in an arbitrary direction. See (13.7a) and Figure 12.16.
ρ	reflectance. See (5.1).
σ	conductivity. See (8.4)
σ^2	variance. See (2.66) and (5.47)
$\sigma^2_{eye-disp}$	perceived external noise variance of the brain-eye-display system. See (13.4 and (13.5c)
σ^2_{eye}	effective internal brain-eye noise variance. See (13.4) and (13.5b)
τ	Transmittance. See (5.1)
τ_{eye}	a time constant of the eye. See (13.9a) and (13.9b).
ψ	a function (usually not known) of dimensionless variables. See (13.24a)
Γ	correlation function. See (4.13)
Φ	flux or powwer. See Table 5.1.
Ω	solid angle. See (5.5) and Figure 5.3.

List of Acronyms

AiTR	Aided target recognizer
APD	Avalanche photodiode
ATR	Automatic target recognizer
BRDF	Bidirectional reflectance distribution function. See Fig. 5.17.
CCD	Charge coupled device
CDF	Cumulative distribution function. See (2.52).
CECOM	U.S. Army Communications & Electronics Command
CTF	Contrast threshold function
DAS	Detector angular subtense
EO	Electro optical 0.4 – 3 μm
FASCODE	Code for calculating atmospheric transmittance. See Section 6.8
FLIR	Forward looking infrared
FOR	Field of regard
FOV	Field of view
IFOV	Instantaneous field of view. See Figsures 3.9 and 8.16.
I^2R	Imaging infrared sensor
IR	Infrared
ISR	Intelligence surveillance and reconnaissance
LED	Light emitting diode
LOWTRAN	Code for calculating atmospheric transmittance. See Section 6.8.
LRG-SWIR	Long range gated short wave infrared
lsf	Line spread function
LSI	Linear shift invariant
LWIR	Long wave infrared 8 – 14 μm
MCT	Mercury cadmium telluride
MODTRAN	Code for calculating atmospheric transmission. See Section 6.8.
MOS	Metal oxide structure
MRC	Minimum resolvable contrast
MRT	Minimum resolvable temperature difference
MRTD	Minimum resolvable temperature difference
MTF	Modulation transfer function
MWIR	Mid wave infrared 3 – 5 μm
NEP	Noise equivalent power. See (8.23), (8.31).
NET	Noise equivalent temperature difference. See (12.5).
NETD	Noise equivalent temperature difference
NIR	Near infrared 0.7 – 1.1 μm
NUC	Non uniformity correction (Section 9.9)
NVESD	Night Vision and Electronic Sensors Directorate
NV-IPM	Night Vision Integrated Performance Model

NVL	Night Vision Lab
oco	optical cut-off
OPL	Optical path length. See (7.3).
OTF	Optical transfer function
pdf	Probability density function. See (2.49)–(2.51)
PMT	Photo-multiplier tube
psf	point spread function
PTF	Phase transfer function
PV	Photovoltaic
ROC	Receiver operating characteristics
ROIC	Read out integrated circuit
SITF	System intensity transfer function. See (8.21) and Figure 8.11.
SNR	Signal to noise ratio. See (8.22).
SR	Spurious response ration. See (12.20)
SWIR	Short wave infrared 1.1 3 μm
TDI	Time delay integration
TOD	Triangle orientation discrimination (Section 12.10)
TRMS	A target acquisition model developed in Germany (Section 12.9)
TTP	Target task performance metric
TTPF	Target transfer probability function

Table of Abbreviations and Concepts

A.S. Aperture stop. See Figures 7.37 and 7.39

Axial ray. A ray that starts and ends on the optical axis. See Figure 7.37.

Chief ray. A ray that originates off the optical axis which goes through the aperture stop. See Figures 7.38 and 7.40

Cold stop. A cold shield with 100% efficiency. See Gigure 7.67

Collimated light. Radiation in which every ray from any given object point can be considered to be parallel to every other ray. See Figure 7.6. This is an ideal which is never realized.

E_{xp}. Exit pupil. See Figures 7.37 and 7.39

E_{xp}. Entrance pupil. See Figure 7.37 and 7.39.

Field stop. Aperture located at an image plane that determines size and shape of an image. See Figure 7.57

Marginal ray. An axial ray that barely makes it through the aperture stop. See Figures 7.37, 7.39 and 7.41.

Nodal points. Of all the rays passing through a lens from an off-axis object point to its corresponding image point, there is always one ray whose direction in image space is the same as the ray in object space. The nodal points are where the extension of these rays intersect the optical axis. See Figures 7.30 and 7.22.

Optical axis. For lenses with spherical surfaces, centers of curvature like on optical axis. See Figure 7.5

Paraxial ray. Axial ray that is always close to the optical axis. See Figure 7.37.

Principal plane. In a lens or lens system, that surface at which the projections of an entering and exiting ray intersect. See Figures 7.13 and 7.17.

Wave front. At a fixed time, the surface over which the optical disturbance is constant. See (4.1).

Table of Operators and Mathematical Functions

$\cup$	Union of two sets. See. (2.58).
$\cap$	Intersection of two sets See (2.57).
$*$	Convolution in one dimension. See (2.9)
$**$	Convolution in two dimensions. See (2.17).
$\star$	Autocorrelation in one dimension. See (2.20a)
$\star\star$	Autocorrelation in two dimensions. See (2.23)
$\mathcal{F}$	Fourier transform of a function in one dimension. See (2.30).
$\mathcal{FF}$	Fourier transform of a function in two dimensions. See (2.31).
J_1	Bessel function of order 1. See Figures 4.11 and 4.12.
step (x)	See Table 2.2
sign (x)	See Table 2.2
rect (x)	See Table 2.2
ramp (x)	See Table 2.2
tri (x)	See Table 2.2
sinc (x)	See Table 2.2
$sinc^2$ (x)	See Table 2.2
Gaus (x)	See Table 2.2
$\delta(x)$	See Table 2.2
$\delta\delta(x)$	See Table 2.2
$\delta_\delta(x)$	See Table 2.2
Comb (x)	See Table 2.2
cyl (r)	See Figure 2.3
Gaus (r)	See Figure 2.3
somb (r)	See Figure 2.3

Index

A

Aberrations, 250–52
 astigmatism, 251
 distortion, 251, 252
 field curvature, 251
 MTF, 251
 spherical, 250–51
Absorption, 168–70
 of a photon, 168
 as quantum process, 168
 scattering versus, 170–71
ACQUIRE model, 380, 384, 529–30
Active/passive detectors, 303–5
Aerosol MTF, 178
Afocal optical system, 237
Aided target recognition, 493–541
Aided target recognizer (AiTR), 4
Aliasing, 63, 334
Analytical parameters, 8–9
Angular magnification, 236–45
 defined, 236
 Keplerian telescope, 237, 238
Apertures
 circular, 249
 numerical, 224–36
 obscured, *f*-number for, 234
 rectangular, 249
 relative, 225
Aperture stops
 chief ray relationship, 224
 defined, 221
 illustrated, 222
 ray tracing for determination, 223
Apodization, 110
Apparent delta T
 defined, 133
 determination techniques, 133–35
 exact transcendental solution technique, 135
 flux-broadband Beer's law product technique, 134–35
 flux-temperature differential technique, 135
 illustrated, 133
 temperature-broadband Beer's law product technique, 134
 temperature-broadband transmission as function of range technique, 134
Arbitrary function, 49
Astigmatic difference, 255
Astigmatism, 251
Astronomical telescopes, 243
Atmospheric MTF, 177–79
 defined, 177
 sampling configuration and, 296
 turbulence, 177
Atmospherics, 163–86, 296
 absorption, 168–70
 components, 163–66
 defined, 163
 exercises, 186
 models, 179–83
 particles, 172
 path radiance, 173
 practical considerations, 183–86
 scattering, 170–73
 structure, 163–66
 transmission, 166
 turbulence, 173–77
 typical transmission, 164
 water vapor, 164–65
Atmospheric scattering coefficient, 141
Atmospheric transmission (MODTRAN), 4, 5
Autocorrelation function, 248
Automatic gain control (AGC), 47–48
Automatic target cueing (ATC) system, 355

Automatic target recognizer (ATR), 4, 355, 369–71
classification, 370
detection, 370
information extraction, 370
performance evaluation curve, 371
preprocessing, 370
process, 370
summary, 371
Avalanche photodiode (APD), 271
Average modulation at optimum phase (AMOP), 391

B

Background-limited infrared photodetection (BLIP), 282, 284, 300–301
Background radiance, 133
Backgrounds
clutter, 150–51
electro-optical systems, 135–41
infrared, 126–35
spatial characteristics, 143–52
spectral emittance, 129
Band-limited functions, 60
Bar target representation, 145–47
illustrated, 147
imaged through turbulence, 184
Beer-Lambert law, 167
Bidirectional reflectance distribution function (BRDF), 142–43
Binomial distribution, 501
Blackbody flux, 444
Blackbody radiation, 126–29
defined, 126
targets, 547
Blur circle, 218, 221, 257
Bolometers, 273–74
Boost filtering, 338–39
Brightness, 326
Buckingham pi theorem, 414
Butterworth filter, 317–18

C

Candelas
defined, 125
per square meter, 417–18
Capacitive transimpedance amplification (CTIA), 330
Carbon dioxide, 165
Cardinal points, 208
Cassegrain telescopes, 244–45
Cathode ray tubes (CRTs), 356, 357–59
defined, 357
example results, 359
illustrated, 357
modulation transfer function (MTF), 406
resolution, 358
sample-and-hold circuit, 359
See also Displays
Charge-coupled devices (CCDs), 275–76
electron-bombarded (EBCCDs), 303–4
electronics, 325–26
high-gain, 298
image intensified (IICCDs), 299
three-phase, 325
Chebychev filter, 317–18
Chief rays, 222
Chromatic aberration, 243
Chromatic lens, 243
Clutter
defined, 150
effect on human target acquisition, 151
modeling, 150–51
modifications, 152
Coatings Engineering Evaluation Program (CREEP), 142
CODE V, 97, 215
Coherence, 77–82
defined, 77
representative lengths, 80
spatial, 79
temporal, 79
Coherent diffraction MTF, 95–96
calculating, 96
circular pupil, 96
as Fourier transform, 95
See also Modulation transfer function (MTF)
Coherent transfer function, 246
Cold shield
defined, 253
efficiency, 446

Cold stop
defined, 253
efficiency, 253
illustrated, 254
Collimated light, 193
Color, 143
Comb, 19, 20
Combat simulation, 529
Comb function, 37–38, 364
1D, 64
2D, 65
Common-module FLIRs, 439
Complex amplitude, 74
Complex functions, 15–16
Concave lenses
behavior summary, 202
defined, 194
image formation by, 195
nomenclature, representation, and behavior, 195
ray-tracing rules for, 197
Conditional probability, 35
Contrast(s), 139–41
defined, 139
example, 141
as function of range in atmosphere, 183
localized enhancement, 336–38
longwave, 152–59
LWIR, 184
midwave, 152–59
MWIR, 184
path radiance, 140
target-to-background, 139
Contrast threshold function (CTF)
alternative forced choice experimental methods, 398
Barten's smooth curves, 399
defined, 11
exercises, 431–33
eye, 369
eye-display contrast threshold model, 424–25
eye-display system, 401–26
in human vision characterization, 385
measured, 385, 386
naked eye, 397–401
system, 427, 475–80
two eyes, 400
Converging beam devices, 291
Convex lenses
behavior summary, 201
defined, 193
first focal point, 205, 206
first principal plane, 205, 206
nomenclature, representation, and behavior, 195
ray-tracing rules for, 196
Convolution, 22–27
defined, 22–24
of delta function, 60
as filtering operation, 24
integral, 50
process, 25
properties, 24
theorem, 30
two-dimensional, in Cartesian coordinates, 26
Cooperative detection, 493
Correlated double sampling (CDS), 330
Correlation, 22–27
cross-correlation, 27
defined, 26
result, 26
theorem, 30
to two dimensions, 27
Correlation function, 78
Cryogenic cooling, 267
Cumulative distribution function (CDF), 34, 498
curves comparison, 508
reduced distribution function versus, 520
for time-limited search, 508
for time unlimited search, 508
Current-mode detector circuits, 309, 311
Cut-on/cut-off frequencies, 430
Cylinder function, 22, 23

D

Dc-restored systems, 317
Decision time, 499

Delta functions, 17, 19, 20
convolution of, 60
in describing arbitrary system input, 49
impulse response as, 52
weighted collection of, 61
Delta T, 117, 131–33
apparent, 133
defined, 131
solution for, 132
target, 147
Depletion region, 269
Detector angular subtense
horizontal and vertical, 286
IFOV and, 287
illustrated, 286
scan velocity and, 314
Detector angular subtense (DAS), 285–87
Detector circuits, 309–13
current-mode, 309, 311
example, 312
photoconductor, 312
voltage-mode, 309, 310
Detector response function, 37
Detectors, 265–305
active/passive, 303–5
background-limited infrared photodetection, 300–301
bolometer, 273–74
charge-coupled device, 275–76
classes of, 265
diamond-shaped, 112
dual-band, 302–3
dwell time, 290
for EO systems, 298–99
equivalent incident power (NEP), 281
exercises, 305–6
flux on, 230–31
infrared, 296–98
infrared arrays, 301–5
laser range-gated, 303–5
modulation transfer function (MTF), 295
noise, 299–300
output signal spectrum, 294
photoconductor, 268–69
photoemissive, 271–72
photon, 265–66
photovoltaic, 269–71
pyroelectric, 274–75
responsivity, 119, 275–78
sampling geometry, 293
scanning configurations and implementations, 287–91
scans across image space, 314
sensitivity, 278–85
SNR characteristics, 278
spatial integration, 292
spectral D^*, 279
square, 112
thermal, 267, 272–75
transfer functions, 292–96
types of, 265–67
voltage, 138
Detector-to-eye filter, 463
Diamond-shaped detector, 112
Differential flux, 132
Differential temperature geometry, 131
Diffraction, 5, 73–116
blur, 256–58
blur circle, 257
coherence and, 77–82
defined, 73
electromagnetic waves and, 74–77
essence of, 73
exercises, 113–16
Fraunhofer, 85–88
Fresnel, 83–85
geometry, 74
psf, 88–91
Diffraction-limited point intensity, 219
Diffraction-limited resolution
thick lens, 257
thin lens, 257
Diffraction modulation transfer function, 73
Diffraction MTF, 91–108
calculation, 96–97, 104
coherent, 95–96
incoherent, 93–95
thin lens, 91–96
Diffraction theory
apodization, 110
applications of, 108–13
detector MTF from Fraunhofer diffraction pattern, 110–13

frequency analysis of optical systems, 108–9
geometric optics, 109
optical image processing, 109
point spread function of distributed aperture, 109
stellar interferometry, 110
Digital filter MTF, 324–25
Dimensional analysis, 401
Diopters, 367
Dirac function, 17
Discrete wavelet transform, 349
Discrimination detection, 493
Displays, 355–65
cathode-ray tubes (CRTs), 357–59
flat-panel, 407, 408
light-emitting diodes (LEDs), 359–61
liquid-crystal displays (LCDs), 361–63
plasma, 363
processing, sampling and, 363–65
reconstructed spectrum after, 365
types of, 356
Distortion, 251, 252
Distributed aperture
horizontal and vertical MTF for, 107
illustrated, 107
Doubling, 47
Dual-band detectors, 302–3
Dual-band infrared systems, 1
Dual *f*-number imager, 256
Dynamic detection. *See* Search
Dynamic MRT (DMRT), 545, 553–55
defined, 554
performance measurements, 554

E

Edge response, 53, 54
Effective focal length, 216–17
defined, 240
telescope-camera optical system, 240
Electromagnetic waves, 74–77
polarization of, 74
representation, 75
Electron Bombarded Active Pixel Sensor (EBAPS), 304
Electron-bombarded CCD (EBCCD), 303–4
Electronics, 309–31
CCDs, 325–26
detector circuits, 309–13
digital filter MTF, 324–25
EO mux MTF, 322–23
exercises, 331
MTF boost filter, 322
noise, 318–21
readout integrated circuits, 328–31
spatial/temporal frequency conversion, 313–15
system component, 356
transfer function, 315–18
uniformity correction or NUC, 326–28
Electro-optical (EO) systems, 3–4, 297–98
contrast, 139–41
defined, 2, 6
detector responsivity, 119
detectors, 298–99
external sources, 136–39
light path, 138
output, 7
performance of, 9–10
primary sources for, 136
resolution, 10
sensitivity, 10
targets and backgrounds, 135–41
target size, 121
turbulence effect on, 176
typical scenario, 6–7
Electro-optical multiplexer (EO mux)
defined, 322
illustrated, 323
MTF, 322–23
Elliptical entrance pupil
horizontal MTF for, 106
illustrated, 105
vertical MTF for, 106
Emissivity, 129–31
defined, 129
measurement techniques, 129–30
representation of, 130
spectral, 130
total, 130, 131

Emittance
defined, 122
differential, 285
exitance and, 122
Entrance pupil
circular, 90, 91, 104
defined, 222
elliptical, 105, 106
racetrack, 108
telescope-camera optical system, 240
Equivalent differential temperature. *See* Delta T
Equivalent incident power (NEP), 278, 281
Equivalent intensity, 280
Error-energy reduction method, 344–45
Euler's equation, 31
Even pulse pair, 19
Exit pupil
defined, 222
human eye, 237
Keplerian telescope, 238, 239
Expected value, 36
External radiation, 6
Extra-wide FOV (XFOV), 254
Eye. *See* Human eye
Eye-brain temporal and spatial integrations, 461, 462
Eye-brain temporal integrations, 461
Eye-display contrast threshold model, 410–26
bandwidth function, 424
Buckingham pi theorem, 414
comparison with existing models, 426
contrast threshold function (CTF), 424–25
dimensional analysis, 414, 420
direct calculations, 433–36
eye transfer function, 424
function definition, 423–26
input parameters, 422, 425
output, 422
program, running, 425–26
temporal noise bandwidth, 423
theory of, 413
time constant of eye, 423
validation of, 410–21
variance internal to eye, 424
Eye-display system, 401–10
equivalent spatial noise bandwidth, 407
eye MTF, 409, 410
Eye photon noise variance, 477
Eyepiece lens, 236
Eye temporal modulation function, 404
Eye time constant, 405

F

Fabrication artifacts, 451
FASCODE, 179
Fast Fourier Transform (FFT), 144
Field curvature, 251
Field of regard (FOR) search
continuously moving sensor over, 511
effectiveness of, 513
model, 511
number of fields of view in, 513
sensor slewing and, 519
single sweep, 512
size of, 511
TDSP search model comparison, 526
time-unlimited, 531–34
validity of, 525
Field of view (FOV), 48, 215–17
extra-wide, 254
illustrated, 214
instantaneous (IFOV), 279, 282
Keplerian telescope, 241
narrow, 186, 217, 254
search, 494–95
source as smaller than, 284
trade-offs, 217
ultra-narrow (UFOV), 254–55
vertical (VFOV), 289
wide, 186, 217, 254
Field stop
added to Keplerian telescope, 242
defined, 215–16
Filtering, image
applications of, 336–39
boost, 338–39
localized contrast enhancement, 336–38
Finite impulse response (FIR) filter, 324
First focal point, 205

First principal focal length, 205
First principal plane, 208
First principal point, 205
First principal surface, 205
Fixed-pattern noise, 451–52
FLIR92 model, 382, 385
Flux
 amount determination, 283
 blackbody, 444
 collection of, 281
 dependence on range, 283
 on detectors, 230–31
f-number, 224–36
 dual, optical systems, 255
 dual imager, 256
 for obscured aperture, 234
 small limit, 230
 as solid angle measure, 236
 symbol representation, 224
 for thick lens in object and image space, 235
 for thin lens in object and image space, 234
 working, 229
Focal lengths
 effective, 216–17
 thick lens, 205, 206
Focal plane
 concept illustration, 198
 defined, 196
Focal point, 204
Focal ratio, 224
Forward-looking infrared (FLIRs), 2, 7
 B-Kits, 382
 common-module, 439
 first-generation, 439
 FLIR92 MRT equation, 382
 generations, 440
 generators of, 7
 MRT model, 384
 NVESD, 383
 second-generation, 440
 third-generation, 253–54, 440
Fourier integral theorem, 30
Fourier transform, 27–30
 coherent transfer function, 246
 defined, 27
 fast (FFT), 144
 inverse, 28
 pairs, 30–32
 properties of, 29–30, 31
 system impulse response, 59
 two-dimensional, 21
Fovea centralis, 366
Fraunhofer diffraction, 85–86
 detector MTF from, 110–13
 irradiance pattern, 111
 from thin lens, 87–88
Fresnel diffraction, 83–85

G

Gamma correction, 326
Gaussian function, 17, 19, 22, 23
Gauss's equation, 201–3
 in finding image location, 212
 image location calculation, 215
 magnification relationship and, 228
General image quality equation (GIQE), 485, 530
 applied to photographic/EO visible systems, 488
 model, 486–89
 results, 489
 use of, 485
Georgia Tech Vision (GTV), 528–29
Ground-resolved distance (GRD), 62–63
Ground-sampled distance (GSD)
 defined, 62
 NIIRS and, 488
 psf and, 62
 as sensor design parameter, 489
 simulated human track data as function of, 490

H

High-pass filter, 317
Historical performance models, 373–95
 exercises, 395
 first FLIRS and, 378–81
 FLIR92, 382, 385
 imager modeling, measurement, field performance, 393

Historical performance models (*cont.*)
improvements for resolution and noise, 381–85
improvement to add sampling, 387–90
introduction to, 373–74
Johnson, 374–76
MRT, 376–78
triangle orientation discrimination (TOD), 392
TRM3, 391–92
Human eye
contrast limitations, 385–87
contrast threshold function (CTF), 369, 397–401
illustrated, 365
modulation transfer function (MTF) of, 367–69
performance of, 401
radii of curvatures, 366
Human-made illuminators, 137
Human perception, 365–69
Human- vision-brain component, 494
Human visual system (HVS)
effective integration time, 477
processing chain, 476
Hybridization, 328

I

Illuminance, 417
Image fusion, 346–50
algorithms, 346–50
applications, 346
Laplacian pyramid, 347–48, 349
objective, 346
quality indices, 350
schematic, 347
wavelet-based methods, 348–50
Image intensified CCDs (IICCDs), 299
Image plane, 196
Image processing, 333–51
filtering, 336–39
fusion, 346–50
sampling theory and, 333–36
summary, 350–51
super-resolution reconstruction, 339–46
Image reconstruction
nonuniform interpolation, 343
process, 342–43
regularized, 343
super-resolution, 339–46
Imagery Interpretability Rating Scale (IIRS), 484, 530
Images, 344
Image space
defined, 225
detector scan across, 314
f-number for thick lens in, 235
f-number for thin lens in, 234, 235
focal point, 193, 194
Imaging
infrared sensors, 2, 3
introduction to, 2–3
of point source on/off axis, 198
process, 64
Imaging systems, 3–4
perfect point, 218
sampled, 64–69
two point sources, 219
wavelength dependencies, 4–6
Impulse response, 50–55
concept, 50
defined, 47
as delta function, 52
edge response, 53, 54
for entire system, 53
example, 51
for perfect optical system, 52
Impulse response function, 45
Incoherent diffraction MTF, 93–95
calculating, 95, 96–97
circular pupil, 96–97
programs for calculating, 97–108
proportional to autocorrelation function, 94
See also Modulation transfer function (MTF)
Independence, 36
Independent dimensionless variables, 414
Index of refraction, 174, 192–93
Index structure parameter, 174
determination, 177

temperature structure parameter relationship, 175
weight, 176
Infinite impulse response (IIR) filter, 324
Infrared CCDs (IRCCDs), 275
Infrared detector arrays, 296, 301–5
active/passive, 303–5
dual-band, 302–3
large-format, 301–2
laser range-gated, 303–5
Infrared detectors, 296–98
InSb, 297
MCT, 296–97
nBn, 297–98
VOx, 298
Infrared imaging system
performance measurements, 545–55
resolution, 549–51
sensitivity, 545–49
Infrared sensors
performance of, 9–10
resolution, 10
sensitivity, 10
target size, 121
typical scenario, 7–8
Infrared targets/backgrounds
apparent delta T, 133
blackbody radiation, 126–29
delta T, 131–33
emissivity, 126–31
exact transcendental solution, 135
flux-broadband Beer's law product, 134–35
flux-temperature differential, 135
target-to-background difference, 125–26
temperature-broadband Beer's law product, 134
temperature-broadband transmission as function of range, 134
wavelengths, 125
Inhomogeneity equivalent temperature difference (IETD), 549
Input irradiance, 53
Instantaneous FOV (IFOV), 279, 282
Intelligence, surveillance, and reconnaissance (ISR) systems, 12, 443
Intersection, 34
Inverse Fourier transform, 28
Irradiance
detector by point source, 231
moonlight, 137
in phasor notation, 78
solar, 136
from sum of two fields for coherent system, 77
Iterative approach, 523–24

J

Johnson detection, 493
Johnson model, 374–76
data relating to resolution, 375
defined, 374
development, 376
experiments, 374–75, 390
target discrimination criteria, 390
Johnson noise, 318–19

K

Keplerian telescope
angular magnification, 237, 238
exit pupil, 238, 239
field of view (FOV), 241
field stop added to, 242
illustrated, 236
Kolmogorov-Obukhov law, 174

L

Ladder filters, 317
Lambertian surface, 123, 124
Laplacian pyramid, 347–48, 349
Large-format arrays, 301–2
Laser range-gated detectors, 303–5
Lensmaker's equation, 199
Light
collimated, 193
representation, 189–91
Light-emitting diodes (LEDs), 356, 359–61
arrays, 323, 360–61
defined, 359–60
illustrated, 360

Light-emitting diodes (LEDs) (*cont.*)
as P-N junction devices, 359
See also Displays
Limiting frequency approach, 470–75
defined, 470
differential temperature, 471
discrimination criteria, 470
probability of discrimination process, 471
target parameter determination, 471
2D MRT, 474
Linear shift-invariant (LSI) systems, 10–11, 45–71
analysis, 45–46
basics of, 48–50
exercises, 69–71
as function of input frequency, 45
input/output behavior, 48
optical, 58
response evaluation, 56
Linear spread function (lsf), 54
Linear systems, 47–48
approach, 10–11
properties of, 47–48
spatial domain, 59
Liquid-crystal displays (LCDs), 356, 361–63
advantages, 362
arrays, 363
defined, 361
as passive devices, 362
See also Displays
Localized contrast enhancement, 336–38
Log-log plots, 412
Long-wave infrared (LWIR) region
carbon dioxide as absorber in, 165
collage example, 157
contrast, 184
defined, 2, 7
human activity discrimination, 483
human scatterplot, 156
images of HMMV and tank, 153
male armed contractor image, 155
rural scatterplot, 157
sandbox images, 158
solar loading effect on, 156
solar reflection, 159
urban scatterplot, 158
use decision, 183
Low-pass filter, 316
LOWTRAN, 380
transmission curves, 180
use of, 179
Lumen, 417

M

Magnetic resonance imaging (MRI), 59
Magnification, 355, 356
Marginal rays, 221
Material impurities, 451
Mathematica search code, 540–41
Mathematics, 15–42
common one-dimensional functions, 17–20
complex functions, 15–16
convolution and correlation, 22–27
examples, 37
exercises, 39–42
Fourier transform, 27–30
probability, 33–37
transform pairs, 30–32
two-dimensional functions, 20–22
MCT avalanche photodiode (MCT-APD), 304–5
Mean, 36
Mean time
detect target with two observers, 536–38
first observer to detect target, 535–36
Meridional plane, 196
Metal organic vapor phase epitaxy (MOVPE), 302
Microdither scanner, 340
Microscanning, 291, 340
Midwave infrared (MWIR) region
camera, MTF for, 551
carbon dioxide as absorber in, 165
collage example, 157
contrast, 184
defined, 2, 7
human activity discrimination, 483
human scatterplot, 156
images of HMMV and tank, 153

male armed contractor image, 155
panoramic imager, 547
rural scatterplot, 157
sandbox images, 158
solar loading effect on, 156
solar reflection, 159
urban scatterplot, 158
use decision, 183
Midwave sensors, 119
Mie scattering, 172
Minimum resolvable contrast (MRC), 11
Minimum resolvable temperature difference (MRT), 376–78, 459–70, 551–53
curve, 381
defined, 376, 459–60, 551
dynamic, 545, 553–55
equation simplification, 378
measurement, 379
measurement process, 552
noise bandwidth, 376–77
parameters, 378
predictions, 381
sampling imaging system measurements, 554
terms in eye integral, 377
2D, 459–70, 553
values, 552
MODTRAN, 380, 473
results, 181
use of, 179
Modulation transfer function (MTF), 11, 457–59
aberration, 251
aerosol, 178
in arbitrary direction, 94
atmospheric, 177–79, 296
boost filter, 322
cathode ray tubes (CRTs), 406
for circular aperture, 249
circular entrance pupil, 104
curve, 460
defined, 47, 57, 549
defining, 92
derivation of, 460–61
detector, 295
diffraction, 91–108
digital filter, 324–25
for elliptical entrance pupil, 106
EO mux, 322–23
expression of, 98
eye, 367–69, 409, 410
horizontal, 458
linear motion, 459
measurement, 460
measurement process, 550
for MWIR camera system, 551
in non-sampled imager characterization, 387
normalizing, 93
optics, 245–49, 295
for racetrack entrance pupil, 108
for rectangular aperture, 249
response, 459
second-generation, 464, 465
as sensor parameter, 460
squeeze approach, 390
system, 441, 457
system horizontal, 468
system versus component, 58–59
turbulence, 177–78
for undersampled imaging systems, 458
vertical, 457
Modulation transfer function compensation (MTFC), 488
Monochromatic sources, 79
Multiple-lens systems, 213–15
Mutually exclusive sets, 35

N

Narrow FOV (NFOV), 186, 217, 254
National Imagery Interpretability Rating Scale (NIIRS), 484–86, 530
definition of, 486
prediction, 486
sample descriptions, 485
Near-infrared (NIR) region, 2
Neoclassical Search Model, 527–28
Newtonian telescopes, 244
Night Vision and Electronic Sensors Directorate (NVESD), 140
Night Vision Integrated Performance Model (NV-IPM), 252, 503–5

Noise
bandwidth, 320–21
detector, 299–300
electronics, 318–21
factor, 321
fixed-pattern, 451–52
Johnson, 318–19
model improvements for, 381–85
1/*f*, 319
pattern, 449
performance, 449–53
photon, 299–300
power spectrum, 321
quantization, 450
readout (floor), 452
reproduction process, 320
reset, 449
shot, 319–20, 449, 450–51
spatial, 381, 548
spatiotemporal, 549
temporal, 381
three-dimensional, 455–57
total current, 452–53
white, 319
Noise equivalent irradiance (NEI), 443, 452, 453–55
exitance of a surface, 453
nomenclature, 454
Noise equivalent temperature difference (NETD), 11, 441, 443
assumptions/approximations, 445
band-averaged detectivity and, 447
computing prior to NUC, 549
concept, 444
defined, 443, 545
derivation, 444
detector, 443
of detector signal, 443
differences in values, 547
estimation of, 545, 546
expression for gross sensitivity issues, 447
generalized form, 445
measurement, 443, 447
overall frame, 443
pixel-level, 546
ratio between electrons and, 479
scanning system parameters, 448
Noise power spectral density, 37
Nonuniform interpolation, 343
Nonuniformity correction (NUC), 328, 549
Normalized correlation function, 78
Normalized optical disturbance, 75
Numerical aperture, 224–36
Nyquist frequency, 334, 335
Nyquist rate, 364, 467

O

Objective lens, 236
Object plane, 196
Object space
defined, 225
f-number for thick lens in, 235
f-number for thin lens in, 234, 235
Object space focal point, 194
Observers, multiple, 514–17
Odd pulse pair, 19
On-chip/off-chip amplifiers, 450
1/*f* noise
defined, 319
knee frequency, 462
One-dimensional functions, 17–20
Optical axis, 193
Optical disturbance amplitude, 95
Optical image processing, 109
Optical materials, 252–53
Optical modulation transfer function, 73
Optical path
illustrated, 189–91
length, 189–91
Optical systems
afocal, 237
frequency analysis of, 108–9
with pupil function, 246
response, 53
telescope-camera, 240
typical, 253–56
vignetting in, 232
Optical transfer function (OTF), 245–46
circular limiting aperture, 38
defined, 57
determination of, 246
Optical transmission, 166

Optics, 189–262
aberrations, 250–52
aperture stop, pupils, rays, 221–24
diffraction blur, 256–58
exercises, 259–62
field of view (FOV), 215–17
f-number and numerical aperture, 224–36
light representation, 189–91
modulation transfer function (MTF), 245–49, 295
multiple-lens systems, 213–15
reflection, 191–93
refraction, 191–93
resolution, 218–21
in sensor systems, 190
spherical mirrors, 203–5
telescopes and angular magnification, 236–45
thick lens, 205–8
thin lens ray-tracing rules, 193–203
vergence, 208–13
Organization, this book, 12

P

Paraxial rays, 221
Parseval's theorem, 30
Path radiance, 140, 173
Pattern noise, 449
Perception time, 499
Perfect point imaging system, 218
Performance
EO noise, 449–53
infrared imaging system, 545–55
measures, 441
minimum resolvable temperature difference, 459–70
modulation transfer function (MTF) and, 457–59
NEI, 453–55
NETD parameter, 443–49
resolution, 442–43
sensitivity, 442–43
system CTF, 475–80
target acquisition with limiting frequency, 470–75
target acquisition with target task performance metric, 480–82
target sets and, 482–84
three-dimensional noise and, 455–57
Persistent ISR systems (P-ISR), 12
Persistent surveillance sensor
coverage, 490
defined, 489
Phase transfer function (PTF), 57–58
defined, 57
system versus component, 58–59
Photoconductor circuits, 312
Photoconductors, 268–69
Photoemissive detectors, 271–72
Photometric units, 124–25
Photometry, 119–25
Photomultiplier tube (PMT), 272
Photon detectors
cryogenic cooling, 267
defined, 265
detection mechanisms, 267
noise, 299–300
photoconductors, 268–69
photoemissive, 271–72
photovoltaic, 269–71
spectral *D**, 279
types of, 265–66
See also Detectors
Photopic efficiencies, 125
Photovoltaic detectors
avalanche photodiode (APD), 271
defined, 269
energy diagram, 270
operation in open-circuit mode, 271
voltage across the junction, 270
PISR, 489
Pixel pitch, 289
Planck's blackbody emission curves, 127
Plane of incidence, 192
Plane waves
direction of travel, 77
propagation, 76
Plasma displays, 363
Point-source intensity example, 123
Point spread function (psf)
amplitude for circular entrance pupil, 90
from circular entrance pupil, 90

Point spread function (psf) (*cont.*)
defined, 5, 47, 73
diffraction, 88–91
distributed aperture, 109
for rectangular aperture, 89
simulated image from circular entrance pupil, 91
square aperture, 113
Polar-to-Cartesian coordinate conversion, 16
Postsample blur transfer function, 67
Presample blur transfer function, 65
Principal plane
defined, 196, 197
extension by practical limits of lens, 200
location of, 199
mirror, 204
Probability, 33–37
conditional, 35
of entire space, 35
use of, 33
Probability density function, 33, 34
for detection time with two observers, 539–40
reduced, for target detection, 524
time-limited search, 507
time-unlimited search, 507
Probability of discrimination
example, 474, 475
process, 471
Programs (incoherent diffraction MTF calculation), 97–108
list of, 98
MTFCyclesPerMmEqHor, 100
MTFCyclesPerMmEqVer, 100
MTFCyclesPerMrEqHor, 101
MTFCyclesPerMrEqVer, 101
MTFNormalizedEqHor, 98, 102
MTFNormalizedEqVer, 99
NMTFCyclesPerMmHor, 100
NMTFCyclesPerMmVer, 100–101
NMTFCyclesPerMrHor, 101
NMTFCyclesPerMrVer, 101–2
NMTFNormalizedHor, 99
NMTFNormalizedVer, 99
Projected area, target, 428
Propagation
plane waves, 76
vergence, 211
Pupil diameter
comparison, 421
as function of field angle/luminance value, 403
Stiles-Crawford correction factor as function of, 403
Pure detection, 493
Pyroelectric detectors, 273–74

Q

Quantization noise, 450
Quasi-monochromatic sources, 79

R

Racetrack entrance pupil, 108
Radiation
blackbody, 126
external, 6, 136
object, spectral distribution and magnitude, 126
photometric units, 124–25
radiometric units, 121–24
reflected, 119
sources, 117–61
Radiometric units, 121–24
Radiometry, 118–25, 229
Raindrops, 172
Ramp function, 17, 18
Random variables, 36
Range performance modeling, 426–31
Rayleigh resolution limit, 219–20
Rayleigh scattering, 172
Ray tracing
in determining aperture stop, 223
to find image location, 212
in showing chief ray/aperture stop relationship, 224
for spherical concave mirror, 204
for spherical convex mirror, 205
through thick lens, 208
Ray-tracing rules
of arbitrary ray, 199

concave lenses, 197
convex lenses, 196
Reaction time, 499
Readout (floor) noise, 452
Readout integrated circuits (ROICs), 328–31
design, 329–30
detector array coupling, 328
manufacturing, 330–31
physical coupling, 329
pixel saturation and, 330
schematic, 328
well capacities, 330
Real images, 200, 201
Receiver operating curve (ROC), 370
Rectangle function, 17, 18, 63
Reflected radiation, 119
Reflectivity
as function of wavelength, 138
at interface, 192
Refraction, law of, 192
Refractive index
in speed of light, 191
thick lens, 207
of vacuum, 192–93
variations, 174
Regions of interest (ROIs), 155
Regularized reconstruction, 343
Re-imaged pupil, 256
Relative aperture, 225
Relative edge response, 487
Reset noise, 449
Resistor-capacitor (RC) circuit, 50
Resolution, 218–21
CRT, 358
defined, 442
definition techniques, 219–21
diffraction-limited, 257, 258
infrared imaging system, 549–51
model improvements for, 381–85
performance, 442–43
Rayleigh limit, 219–20
sensitivity relationship, 443
spatial sampling and, 62–64
Resolved source, 280
Responsivity
curve, 277
detectors, 275–78
measurement, 276
Reticle, 242
Rotating-drum scanners, 291

S

Sagittal plane, 196
Sampled imaging systems, 64–69
output, 68
presample blur, 65, 66
RSS, 68
sampled signal and display transfer, 67
sampling output, 66
Sampled spectrum, 61
Sampling
detector geometry, 293
display processing and, 363–65
function, 62
goal, 333
model improvement to add, 387–90
performance loss, 389
theory, 333–36
Scanning systems
configurations, 287–91
microscanning, 291
overscan ratio, 289, 290
parallel, 288, 289
parameters, 289
serial, 288
vertical MRT spatial integrals, 463
Scan velocity, 462
Scattering, 170–73
absorption versus, 170–71
categories of models, 171
coefficients, 141, 171
Mie, 172
by particles, 171
process, 170
raindrops, 172
Rayleigh, 172
Sclera, 366
Scotopic efficiencies, 125
Search, 493–541
combat simulation and war games, 529
defined, 493
exercises, 530–31

Search (*cont.*)
field of regard, 510–13
FOR, 494–95
FOV, 494–95
Georgia Tech Vision (GTV), 528–29
independent, 517–18
Mathematica code, 540–41
multiple observers, 514–17
neoclassical, 527–28
parameter estimation techniques, 499–503
parameters, 503–5
shared knowledge, 514–17
single sensor, 514–17
step-stare, 526
theory, 495–99
time-dependent parameters, 518–27
time-limited, 505–10
time-unlimited, 507, 508, 514–17
two sensors, 517–18
Second focal point, 205
Second-generation MRT
FLIR example, 469
parameters, 464
spatial pre-filter rollup, 465
Second principal focal length, 206
Second principal focal point, 205
Second principal plane, 205, 206
Second principal point, 205–6
Second principal surface, 205
Sensitivity, 9
commercial camcorders, 455
considerations, 142–43
detector, 278–85
infrared imaging system, 545–49
performance, 442–43
resolution as conflicting parameter, 10
resolution relationship, 443
Sensors
optics in, 190
slewing, 519
turbulence limited, 184
Serial scan ratio, 289
Shared information
detection times and probabilities, 534–40
mean time for first observer, 535–36
mean time to detect target, 536–38
pdf for detection time, 539–40
useful mathematical result, 534–35
Shift invariance, 48, 50
Shift-invariant discrete wavelet transform (SIDWT), 349–50
Shift theorem, 29
Short exposure time, 177
Short-wave infrared (SWIR) region, 2
Shot noise, 319–20, 449, 450–51
Signal-to-noise ratio (SNR), 285
Signum function, 18
Similarity theorem, 29
Sinc function, 18, 364
Snell's law, 192, 198
Solar irradiance, 136
Solar loading, 156
Solar reflection contribution, 159
Solid angle, 121
Sombrero function, 22, 23
Sparrow criterion, 220
Spatial coherence
concepts, 81
defined, 79
length, 81, 82
source separation and, 81, 82
Spatial domain linear systems, 59
Spatial exitance, 144
Spatial eye integrals, 468
Spatial frequency
content, 144
conversion, 313–15
domain, 59
Spatial post-filter transfer, 467
Spatial resolution, 9, 10
Spatial sampling, 59–62
accurate representation, 59
resolution and, 62–64
Whittaker-Shannon sampling theorem, 60, 62
Spatiotemporal noise, 549
Spectral emittance
defined, 126
of target or background, 129
Spectral emitters, 127
Spherical aberrations, 250–51

Spherical mirrors, 203–5
Spurious response
 calculation in cycles per millimeter, 389–90
 defined, 388
Square detector, 112
Standard deviation, 36, 37, 501, 502
Star brightness, 137
Staring systems, 287, 449, 2980
Stellar interferometry, 110
Step function, 17, 18
Stephan-Boltzmann constant, 127
Step-stare search, 526
Steradian, 121, 122
Stiles-Crawford correction factor, 403
Strehl ratio, 245
Subpixel shift estimation, 340–42
Super-resolution image reconstruction, 339–46
 acronym, 344
 defined, 339
 dynamic imagery baseline, 346
 error-energy reduction method, 344–45
 example and performance estimates, 343–46
 experimental results, 345
 human perceptual, 346
 image acquisition, 340
 image reconstruction, 342–43
 as linear inverse problem, 339
 nonuniform interpolation, 343
 overview, 339
 regularized, 343
 subpixel shift estimation, 340–42
Synthetic aperture radar (SAR), 59
System CTF, 427, 475–80
 calculation parameters, 478
 demonstration, 478
 equation, 475–76
 noise example values, 479
 plotted at different frequencies, 480
System intensity transfer functions (SITFs), 11, 278
System magnification, 355, 356
System noise bandwidth, 462
System transfer function, 62

T

Tangential plane, 196
Target acquisition
 clutter effect on, 151
 with limiting frequency approach, 470–75
 probability in step-stare search, 526
 probability of discrimination, 496
 with target task performance metric, 480–82
Target Acquisition Draw Methodology (TADM), 529
Target delta T, 147
Target detection
 first observer, mean time, 535–36
 pdf for, 524
 with two observers with sensors, 536–38
Target models, 152
Target radiance, 133
Targets
 bar target representation, 145–47
 blackbody, 547
 characteristic dimension, 429
 characteristics, simulation of, 151–52
 characteristics summary, 148–50
 contrast, 429
 discimrination tasks, 149–50
 discrimination definitions, 149
 electro-optical systems, 135–41
 identification range, 184
 infrared, 126–35
 projected area of, 428–29
 spatial characteristics, 143–52
 spectral emittance, 129
 transfer performance, 429
 vehicle, characteristics, 148
Target sets, 482–84
 calibration constants, 482
 parameters, 484
Target task performance (TTP), 1, 426–31
 application of, 390
 cut-on/cut-off frequencies, 430
 improvements prior to, 390–91
 with limiting frequency approach, 481

Target task performance (TTP) (*cont.*)
metric development, 475
target acquisition with, 480–82
value reduction, 481
Target transfer probability function (TTPF), 1, 146, 379
illustrated, 473
probability of detection determination with, 472
Telescopes, 236–45
astronomical, 243
Cassegrain, 244–45
field of view (FOV), 241
Keplerian, 236, 237–39
Newtonian, 244
terrestrial, 243
Temperature
differential, geometry, 131
as function of altitude, 165
structure parameter, 174
thermal detector operation, 273
See also Minimum resolvable temperature difference (MRT)
Temporal coherence
concepts, 81
defined, 79
length, 80
Temporal frequency conversion, 313–15
Temporal post-filter transfer, 466
Terrestrial telescopes, 243
Thermal detectors
bolometer, 273–74
defined, 265
operation temperature, 273
pyroelectric, 274–75
response time, 273
types of, 267
See also Detectors
Thermal reflection contribution, 159
Thick lens
diffraction-limited resolution, 258
focal lengths, 205, 206
modeling, 205–8
nodal points, 207
principal planes, 205, 206
ray tracing through, 208
refractive index, 207
Thin lens
diffraction-limited resolution, 257
diffraction MTF, 91–96
Fraunhofer diffraction from, 87–88
Gauss's equation, 201–3
optical system diffraction PSF, 88–91
ray-tracing rules, 193–203
Three-dimensional noise, 455–57
components, 456
coordinates, 455
estimates based on historical measurements, 457
use of, 455
See also Noise
Time-dependent search parameters (TDSP), 518–27
calculation, 527, 528
comparison of, 525
Mathematica search code, 540–41
search model, 518–27
FOR search model comparison, 526
search theory, 519–20
Time-limited search, 505–10
CDF comparison with delay times, 510
CDF for, 508
defined, 506
pdfs comparison with delay times, 509
probability density function for, 507
Time-unlimited search
CDF for, 508
FOR, 531–34
probability density function for, 507
Total noise current, 452–53
Transfer function, 55–58
coherent, 246, 247
defined, 55
detector, 292
diffraction modulation, 73
electronics, 315
example, 57
incoherent, 247
noncoherent optical (OTF), 38, 57, 245–46
optical modulation, 73
phase (PTF), 57–58
postsample blur, 67
presample blur, 65

system, 59, 62
system intensity, 278
weighting determination, 56
See also Modulation transfer function (MTF)
Transform pairs, 30–32
list of, 32
size of, 32
See also Fourier transform
Transmission
atmospheric, 166–68
of atmospheric gases, 169
high-resolution, 170
optical, 166
of radiation, 166
Transverse electromagnetic (TEM) waves, 74
Triangle function, 17, 18, 21
Triangle orientation discrimination (TOD), 343, 392
TRM3 model, 391–92
Trolands, 417–18
Turbulence, 173–77
bar targets imaged through, 184
defined, 173–74
effect on E/O systems, 176
Turbulence MTF
atmospheric, 177
illustrated, 178
short-exposure-time, 177–78
2D MRT, 459–70
curve, 553
determination, 469
illustrated, 470
for limiting frequency approach, 474
Two-dimensional functions, 20–22

U

Ultra-narrow FOV (UFOV), 254–55
Uniformity correction, 326–27
Union, 34
Unresolved source, 283
Unsharp masking, 337

V

Variance, 36
Vehicle target characteristics, 148
Vergence
defined, 208
examples of, 209, 210
in finding image location, 212
propagation, 211
rules, 209
surface power and, 211
as wavefront curvature, 209
Vertex plane, 237
Vertical FOV (VFOV), 289
Vignetting, 232
Virtual images, 200, 201
Voltage-mode detector circuits, 309, 310

W

War games, 529
Wavelength dependencies, 4–6
Wavelet-based methods, 348–50
Wavelet transform sensor fusion, 349
Weighting factor, 175
Well capacities, 330
White noise, 319
Whittaker-Shannon sampling theorem, 61
Wide FOV (WFOV), 186, 217, 254
Wien's law, 126–27, 128

Z

Zeemax, 215